

Министерство образования и науки Российской Федерации
ФГБОУ ВО «Уральский государственный педагогический университет»
Институт математики, физики, информатики и технологий
Кафедра физики и математического моделирования

**Информационные технологии для бизнес-анализа в
коммерческой организации**

Выпускная квалификационная работа

Квалификационная работа
допущена к защите
Зав.кафедрой ФиММ
д.ф-м.н., профессор
Сидоров Валерий Евгеньевич

Исполнитель:
Семенова Екатерина Витальевна
обучающийся БЭ-51з группы

подпись

дата

подпись

Руководитель:
Кощеева Елена Сергеевна,
к.п.н., доцент кафедры ФиММ

подпись

Екатеринбург 2017

Оглавление

Введение	3
Глава 1. Основные концепции бизнес анализа	5
1.1. Современный бизнес анализ	5
1.2. Современные аналитические платформы обработки данных	16
Глава 2. Информационные технологии для бизнес-анализа.....	33
2.1. Возможности аналитической платформы Deductor	33
2.2. Бизнес-анализ деятельности торговой компании в аналитической платформе Deductor	43
2.3. Ассоциативные правила в стимулировании розничных продаж.....	58
Заключение	67
Список литературы	68

Введение

В условиях современной экономики для каждого предприятия становится актуальным повысить конкурентоспособность и сохранять ее на должном уровне. С этой задачей помогут справиться системы бизнес-аналитики, которые позволяют руководителям принимать обоснованные управленческие решения при существенной экономии времени на поиск и анализ необходимой информации.

Целью дипломной работы является обзор популярных систем бизнес-аналитики, выбор наиболее подходящей BI (Business intelligence) для организации и разработки готовых сценариев для проведения бизнес-анализа.

Для достижения цели в рамках дипломной работы необходимо решить основные задачи:

- изучить основные концепции бизнес анализа;
- изучить современный рынок систем бизнес-анализа;
 - выбрать аналитическую платформу, которая учитывает специфику отрасли торговли;
- разработать сценарии для проведения бизнес-анализа на предприятии.

Практическая значимость работы заключается в том, что результаты, полученные в ней, могут быть реализованы в деятельности компании ООО «Автолидер».

Выпускная квалификационная работа состоит из введения, двух глав и заключения.

Во введении рассматривается актуальность темы исследования, определена цель и сформулированы задачи дипломной работы.

В первой главе описываются основные концепции современного бизнес-анализа, рассматриваются наиболее часто используемые технологии интеллектуального анализа, современные аналитические платформы обработки данных.

Во второй главе детально описываются возможности аналитической платформы Deductor, ее назначение, организационная структура, состав и назначение основных модулей платформы. Так же проводится бизнес-анализ деятельности торговой компании в системе Deductor.

В заключении изложены основные результаты работ, обеспечившие достижение цели выпускной квалификационной работы и решение поставленных задач.

Глава 1. Основные концепции бизнес анализа

1.1. Современный бизнес анализ

Современный бизнес анализ считается одним из самых эффективных инструментов для управления компанией, при помощи которого можно получить реальное представление об общем положении в компании и оценить перспективу развития в долгосрочном периоде.

Бизнес-анализ, это набор задач и техник, используемых как связующее между участниками и заинтересованными лицами для понимания структуры, правил и операций в рамках организации, так и для рекомендации решений, позволяющих данной организации достичь ее целей [27].

Есть несколько вариантов определения понятия «бизнес-анализ». В узких областях знаний и в различных отраслях этот термин может интерпретироваться по-разному.

В менеджменте и экономических дисциплинах категория «бизнес-анализ» определяется как совокупность механизмов и путей выявления «узких мест» в компаниях, точек роста, определения проблем организации и поиска решений этих вопросов путем последующей перестройки некоторых процессов [21]. В такой трактовке анализ нужен бизнесу для:

- создания общей картины состояния бизнес-процессов в организации;
- привлечения специалистов для совместного поиска путей выхода из проблемных ситуаций;
- реинжиниринга бизнес-процессов с целью достижения эффективности деятельности;
- формирования универсальных подходов к выполнению определенных задач в организации;

- анализа эффективности и других показателей работы над группой взаимосвязанных задач;
- ввода инновационных технологий, компьютерных программ и прочего.

Бизнес-анализ объединяет:

- рыночные условия;
- материально-производственную базу;
- инновационные производственные, компьютерные, рекламные и другие механизмы;
- задействованные основные и оборотные средства, денежные гарантии и риски;
- занимаемую нишу фирмы в рыночном окружении и возможные шаги конкурентов;
- общественную, финансовую и политическую среду, в которой существует фирма.

Результатом должна являться разработка эффективной стратегии, направленной на максимизацию прибыли в краткосрочной и долгосрочной перспективе и, в конечном итоге, на исполнение миссии организации [23].

Технологии бизнес-анализа

Data mining

Развитие методов записи и хранения данных привело к бурному росту объемов собираемой и анализируемой информации. Объемы данных настолько внушительны, что человеку просто не по силам проанализировать их самостоятельно, хотя необходимость проведения такого анализа вполне очевидна, ведь в этих "сырых" данных заключены знания, которые могут быть использованы при принятии решений. Для того чтобы провести автоматический анализ данных, используется Data Mining.

Data Mining – это процесс обнаружения в "сырых" данных ранее неизвестных нетривиальных практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности [16]. Data Mining является одним из шагов Knowledge Discovery in Databases.

Информация, найденная в процессе применения методов Data Mining, должна быть нетривиальной и ранее неизвестной, например, средние продажи не являются таковыми. Знания должны описывать новые связи между свойствами, предсказывать значения одних признаков на основе других и т.д. Найденные знания должны быть применимы и на новых данных с некоторой степенью достоверности [8]. Полезность заключается в том, что эти знания могут приносить определенную выгоду при их применении. Например, проще всего воспринимаются человеком логические конструкции "если ... то ...". Более того, такие правила могут быть использованы в различных СУБД в качестве SQL-запросов. В случае, когда извлеченные знания непрозрачны для пользователя, должны существовать методы постобработки, позволяющие привести их к интерпретируемому виду.

Алгоритмы, используемые в Data Mining, требуют большого количества вычислений. Раньше это являлось сдерживающим фактором широкого практического применения Data Mining, однако сегодняшний рост производительности современных процессоров снял остроту этой проблемы. Теперь за приемлемое время можно провести качественный анализ сотен тысяч и миллионов записей.

Задачи, решаемые методами Data Mining, делятся на следующие категории:

Классификация – это наиболее распространенная задача, заключающаяся в разбиении объектов на заранее известные классы. Каждый класс обладает определенным набором признаков и по этим признакам любой объект может быть отнесен к одному из существующих классов [12].

Регрессия – заключается в поиске функции задаваемых параметров, которая определяла бы спектр значений искомой величины (предсказание погоды исходя из имеющихся исторических наборов данных).

Кластеризация – это группировка объектов (наблюдений, событий) на основе данных (свойств), описывающих сущность этих объектов. Объекты внутри кластера должны быть "похожими" друг на друга и отличаться от объектов, вошедших в другие кластеры. Чем больше похожи объекты внутри кластера и чем больше отличий между кластерами, тем точнее кластеризация.

Ассоциация – выявление закономерностей между связанными событиями. Примером такой закономерности служит правило, указывающее, что из события X следует событие Y. Такие правила называются ассоциативными. Анализ совершаемых покупок и выявление товаров, часто продаваемых вместе – типичный пример задачи. Иногда ее так и называют анализом рыночной корзины (market basket analysis).

Последовательность – нахождение временной закономерности между событиями, т.е. обнаружение зависимости, что если произойдет событие X, то спустя заданное время произойдет событие Y (например, после покупки автомобиля, автовладелец оформляет страховку, приобретает аптечку, огнетушитель и аварийный знак).

Анализ отклонений – обнаружение в наборе данных наиболее отличающиеся, нехарактерные значения.

Решение большинства проблем бизнес-анализа сводится к той или иной задаче Data Mining или к их комбинации. Например, оценка рисков – это решение задачи регрессии или классификации, сегментация рынка – кластеризация, стимулирование спроса – ассоциативные правила. Фактически, задачи Data Mining являются элементами, из которых можно собрать решение подавляющего большинства реальных бизнес задач.

Для решения вышеописанных задач используются различные методы и алгоритмы Data Mining. Ввиду того, что Data Mining развивалась и развивается на стыке таких дисциплин, как статистика, теория информации, машинное обучение, теория баз данных, вполне закономерно, что большинство алгоритмов и методов Data Mining были разработаны на основе различных методов из этих дисциплин [2]. Большую популярность получили следующие алгоритмы Data Mining: деревья решений; алгоритмы кластеризации; нейронные сети; алгоритмы обнаружения ассоциативных связей между событиями.

KDD

Knowledge Discovery in Databases (KDD) – это процесс поиска полезных знаний в "сырых" данных. KDD включает в себя вопросы: подготовки данных, выбора информативных признаков, очистки данных, применения методов Data Mining (DM), постобработки данных и интерпретации полученных результатов [4]. Безусловно, "сердцем" всего этого процесса являются методы DM, позволяющие обнаруживать знания.

Этими знаниями могут быть правила, описывающие связи между свойствами данных (деревья решений), часто встречающиеся шаблоны (ассоциативные правила), а также результаты классификации (нейронные сети) и кластеризации данных (карты Кохонена) и т.д.

Процесс Knowledge Discovery in Databases, состоит из следующих шагов:

1. **Подготовка исходного набора данных.** Этот этап заключается в создании набора данных, в том числе из различных источников, выбора обучающей выборки и т.д. Для этого должны существовать развитые инструменты доступа к различным источникам данных. Желательно иметь поддержку работы с хранилищами данных и наличие семантического слоя, позволяющего использовать для

подготовки исходных данных не технические термины, а бизнес понятия.

2. **Предобработка данных.** Для того чтобы эффективно применять методы Data Mining, следует обратить внимание на вопросы предобработки данных. Данные могут содержать пропуски, шумы, аномальные значения и т.д. Кроме того, данные могут быть избыточны, недостаточны и т.д. [30]. В некоторых задачах требуется дополнить данные некоторой априорной информацией. Не верно полагать, что если подать данные на вход системы в существующем виде, то на выходе получим полезные знания. Данные должны быть качественны и корректны с точки зрения используемого метода DM. Поэтому первый этап KDD заключается в предобработке данных. Более того, иногда размерность исходного пространства может быть очень большой, и тогда желательно применять специальные алгоритмы понижения размерности. Это как отбор значимых признаков, так и отображение данных в пространство меньшей размерности.

3. **Трансформация, нормализация данных.** Этот шаг необходим для приведения информации к пригодному для последующего анализа виду. Для чего нужно проделать такие операции, как приведение типов, квантование, приведение к "скользящему окну" и прочее. Кроме того, некоторые методы анализа, которые требуют, чтобы исходные данные были в каком-то определенном виде. Нейронные сети, скажем, работают только с числовыми данными, причем они должны быть нормализованы.

4. **Data Mining.** На этом шаге применяются различные алгоритмы для нахождения знаний. Это нейронные сети, деревья решений, алгоритмы кластеризации, установления ассоциаций и т.д.

5. **Постобработка данных.** Интерпретация результатов и применение полученных знаний в бизнес приложениях.

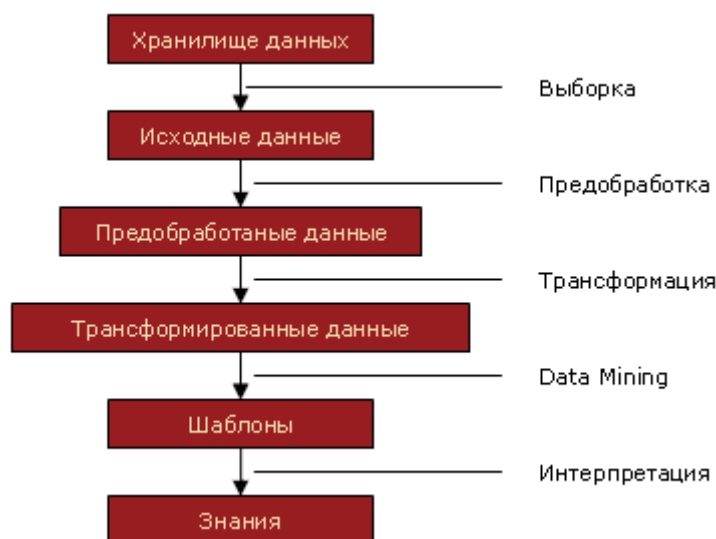


Рис. 1. Процесс Knowledge Discovery in Databases

Knowledge Discovery in Databases не задает набор методов обработки или пригодные для анализа алгоритмы, он определяет последовательность действий, которую необходимо выполнить для того, чтобы из исходных данных получить знания [13]. Данный подход универсальный и не зависит от предметной области, что является его несомненным достоинством.

OLAP

Механизм OLAP является на сегодня одним из популярных методов анализа данных. Есть два основных подхода к решению этой задачи. Первый из них называется Multidimensional OLAP (MOLAP) – реализация механизма при помощи многомерной базы данных на стороне сервера, а второй Relational OLAP (ROLAP) – построение кубов "на лету" на основе SQL запросов к реляционной СУБД [5]. Каждый из этих подходов имеет свои плюсы и минусы. Здесь описана реализация ядра настольного ROLAP модуля.

Такая задача возникла после применения ROLAP системы, построенной на основе компонентов Decision Cube, входящих в состав Borland Delphi [11]. К сожалению, использование этого набора компонент показало низкую производительность на больших объемах данных. Остроту

этой проблемы можно снизить, стараясь отсечь как можно больше данных перед подачей их для построения кубов. Но этого не всегда бывает достаточно.

Схема работы

Общую схему работы настольной OLAP системы можно представить следующим образом:

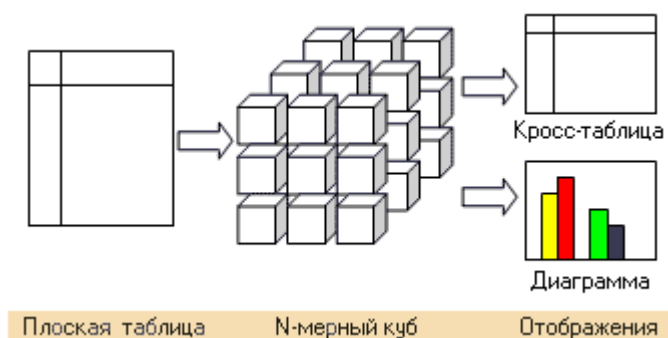


Рис. 2. Общая схема OLAP

Алгоритм работы следующий:

- Получение данных в виде плоской таблицы или результата выполнения SQL запроса.
- Кэширование данных и преобразование их к многомерному кубу.
- Отображение построенного куба при помощи кросс-таблицы или диаграммы и т.п. В общем случае, к одному кубу может быть подключено произвольное количество отображений [33].

Рассмотрим, как подобная система может быть устроена внутри. Отображения, используемые в OLAP системах, чаще всего бывают двух видов – кросс-таблицы и диаграммы. Рассмотрим кросс-таблицу, которая является основным и наиболее распространенным способом отображения куба.

Кросс-таблица

На приведенном ниже рисунке, желтым цветом отображены строки и столбцы, содержащие агрегированные результаты, светло-серым цветом

отмечены ячейки, в которые попадают факты и красным ячейки, содержащие данные размерностей.

Рис. 3. Кросс-таблица

Таким образом, таблицу можно разделить на следующие элементы:

Заголовки столбцов						
Заголовки строк						
Матрица с фактами						

Рис. 4. Фрагмент кросс-таблицы

Заполняя матрицу с фактами, необходимо действовать следующим образом:

1. На основании данных об измерениях определить координаты добавляемого элемента в матрице.
2. Определить координаты столбцов и строк итогов, на которые влияет добавляемый элемент.

3. Добавить элемент в матрицу и соответствующие столбцы и строки итогов.

При этом нужно отметить то, что полученная матрица будет сильно разреженной, почему ее организация в виде двумерного массива (вариант, лежащий на поверхности) не только нерациональна, но, скорее всего, и невозможна в связи с большой размерностью этой матрицы, для хранения которой не хватит никакого объема оперативной памяти. Допустим, куб содержит информацию о продажах за летний сезон, и если в нем будет всего 3 измерения – Клиенты (190), Товары (150) и Дата (92), то мы получим матрицу фактов следующих размеров:

$$\text{Кол-во элементов} = 190 \times 150 \times 92 = 2\,622\,000$$

При том, что заполненных элементов в матрице может быть всего несколько тысяч. Причем, чем больше количество измерений, тем более разреженной будет матрица.

Поэтому, для работы с этой матрицей нужно применить специальные механизмы работы с разреженными матрицами. Возможны различные варианты организации разреженной матрицы.

Рассмотрим теперь, как можно определить координаты факта, зная соответствующие ему измерения. Для этого рассмотрим подробнее структуру заголовка:

Изм. 1			Изм. 2			Итог
Изм. 1	Изм. 2	Итог	Изм. 1	Изм. 2	Итог	

Рис. 5. Заголовок таблицы

При этом можно легко найти способ определения номеров соответствующей ячейки и итогов, в которые она попадает. Здесь можно предложить несколько подходов. Один из них – это использование дерева для поиска соответствующих ячеек. Это дерево может быть построено при проходе по выборке. Кроме того, можно легко определить аналитическую рекуррентную формулу для вычисления требуемой координаты.

Подготовка данных

Данные, хранящиеся в таблице необходимо преобразовать для их использования. Так, в целях повышения производительности при построении гиперкуба, желательно находить уникальные элементы, хранящиеся в столбцах, являющихся измерениями куба. Кроме того, можно производить предварительное агрегирование фактов для записей, имеющих одинаковые значения размерностей. Здесь важны уникальные значения, имеющиеся в полях измерений. Тогда для их хранения можно предложить следующую структуру:

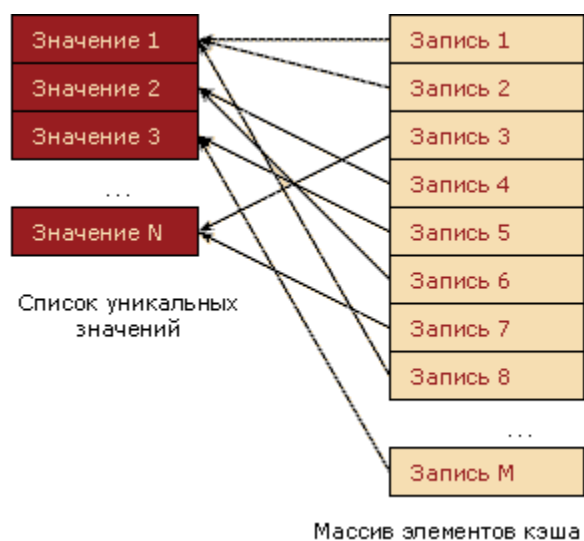


Рис. 6. Структура хранения

При использовании такой структуры значительно снижается потребность в памяти [22]. Что довольно актуально, т.к. для увеличения скорости работы желательно хранить данные в оперативной памяти. Кроме того, хранить можно только массив элементов, а их значения выгружать на диск, так как они будут требоваться только при выводе кросс-таблицы.

1.2. Современные аналитические платформы обработки данных

Система бизнес-анализа класса business-intelligence это:

1. Программа, которая включает в себя схемы, таблицы, таблицы, дашборды и отчеты, индивидуально отрегулирована для комфортного осуществления стратегического и операционного бизнес-анализа [17].
2. Необходимое приложение для принятия результативных решений.
3. Эффективный инструмент для повышения продуктивного руководства, сокращения финансовых трат и увеличения прибыли.
4. Возможность иметь доступ к подробным данным о проекте в любой точке планеты [26].

Программа бизнес-анализа систематически осуществляет сбор информации из всех существующих отделов компании и в краткой форме передает самые важные данные, сверенные с установленными показателями, ответственному специалисту. В это же время приложение позволяет вывести информацию любого уровня, в том числе первичную документацию. Система бизнес-анализа необходима для того, чтобы в режиме онлайн наблюдать за результатами работы организации и в случае необходимости незамедлительно действовать, если процессы проекта отклоняются от поставленных показателей или установленного плана.

В получении важных данных из первичной необработанной информации нуждаются как серьезные компании, так и организации среднего и малого бизнеса. Системы бизнес-анализа многофункциональны, поэтому они подходят для решения многих вопросов во всех сферах деятельности.

Приложение бизнес-анализа существует самостоятельно и неспособно решать или же заменять другие системы, но имеет доступ к информации из

многих источников – от документов формата Excel и решений, прописанных самостоятельно, до CRM, ERP и WMS-систем [25].

Системы бизнес-анализа способны осуществить подключение к любой из существующих в организации IT-систем и загрузить в себя все имеющиеся данные. Загрузка целиком проходит однажды, а затем можно лишь обновлять данные, это способствует моментальному выводу результата бизнес-анализа.

Доступом к аналитической информации в приложении бизнес-анализа обладают учредители и генеральный директор, а также аналитики и менеджеры компании. Графический интерфейс системы бизнес-анализа, как правило, прост и понятен, дополнительных навыков при работе не требует. Внедрение системы бизнес-анализа не требует перемен в рабочем процессе. Его внедрение осуществляется за короткий период практически без отрыва персонала от работы. Достоинства системы бизнес-анализа легко подсчитать, так как эффект от ее использования появляется мгновенно и легко вычисляется в денежном эквиваленте.

Типичная система бизнес-анализа обладает такими показателями (по данным журнала «Коммерческий директор» от 15.03.2016):

- увеличение производительности труда – 34 %;
- повышение процента удовлетворенности заказчиков компании – 96 %;
- возвращение вложенных финансовых средств – 186 %;
- увеличение прибыли – 16 %;
- рост финансовых инвестиций – 23 %;
- уменьшение затрат на эксплуатацию техники – 20 %.

Системы бизнес аналитики

Сегодня существует большое количество BI решений с различной степенью готовности и целым комплексом реализованных функций как крупных мировых вендоров, таких как Microsoft или IBM, так и небольших

компаний, специализирующихся на разработке BI платформ. Согласно Gartner, сегодня перед разработчиками BI платформ стоит задача удовлетворить потребности в организации доступа и инструментах обработки аналитических данных как простых бизнес-пользователей, так и продвинутых IT специалистов [18]. Для этого разработчики стараются сделать интерфейсы своих систем интуитивно понятными всем пользователям, при этом постоянно расширяя спектр аналитических возможностей системы для извлечения, обработки и загрузки все большего количества и типов данных, совершенствуя механизмы построения сложных прогностических и аналитических моделей.

Трендом последних двух лет среди разработчиков стало построение своих BI платформ на основе технологий Data Discovery [28]. Решения типа Data Discovery предлагают пользователям интерактивный графический интерфейс, базирующийся на архитектуре in-memory (в оперативной памяти), что обеспечивает простоту и быстроту работы системы, а также позволяет работать с данными в режиме drag-and-drop. По мере накопления данных, как собственных, так и Big data (“Больших данных”), увеличения их количества и совершенствования средств их обработки растет спрос на развитие и повышение точности инструментов прогнозирования. Возможности построения точных прогнозов на основе большого объема аналитических данных из различных источников позволяют повысить эффективность планирования ресурсов и снизить риски для организаций.

С развитием мобильных технологий и их широким распространением возник спрос на возможность просмотра аналитических данных в любом месте и в любое время с мобильных устройств. Основные разработчики BI платформ вовремя отреагировали на запрос пользователей, выпустив в дополнение к своим основным BI продуктам приложения для смартфонов и планшетов для популярных мобильных платформ.

Исходя из того, что на рынке представлено более 20 BI-платформ, были отобраны семь наиболее популярных систем для подробного рассмотрения:

1. QlikView
2. Tableau
3. Prognosz Platform
4. IBM Cognos
5. Microsoft Power BI
6. Tibco Spotfire
7. Deductor Studio Academic

Далее опишем кратко каждую из рассмотренных платформ.

QlikView

Компания QlikTech со своим BI решением QlikView является пионером в области разработки платформ для бизнес-аналитики и на сегодняшний день вместе с Tableau делит первое место на рынке. QlikTech имеет офисы по всему миру, около 28 000 клиентов в более чем 100 странах мира.

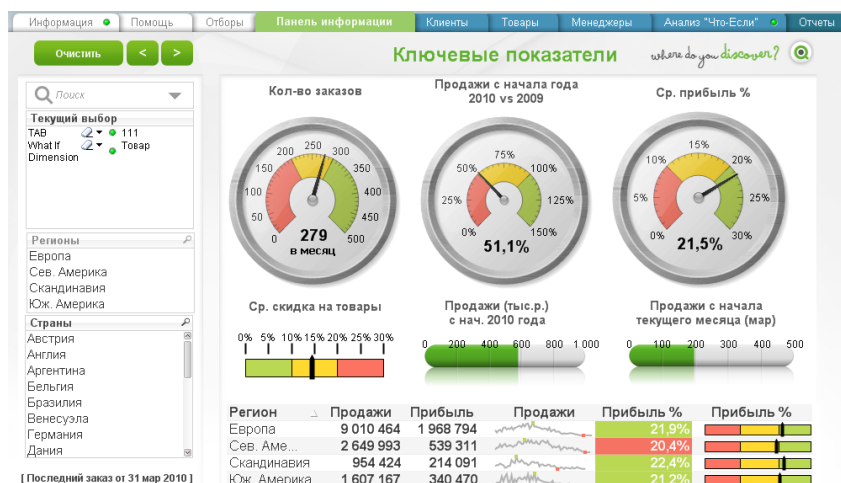


Рис.7. Рабочее окно QlikView.

QlikView – это самостоятельная BI платформа, использующая ассоциативную модель данных, базирующаяся на системе обработки запросов в оперативной памяти (in-memory) [35]. Запатентованный механизм QlikView для обработки данных в оперативной памяти способен выполнять

десятикратное сжатие обычных данных; благодаря этому на один сервер, оснащенный 256 Гбайт оперативной памяти, можно загрузить более 2 Тбайт несжатых данных. Это позволяет работать с миллиардами строк, обеспечивая при этом время отклика в несколько секунд. В QlikView встроен собственный ETL-обработчик, поэтому можно грузить и обрабатывать любые данные. Вместо используемых в других платформах OLAP-кубов, QlikView предлагает свой инструмент бизнес-аналитики (Business Discovery), более эффективный, по их мнению, чем OLAP.

Пользователи по всему миру выбирают QlikView за хорошо проработанные дашборды, интерактивную визуализацию, геопространственную аналитику, возможность одновременного использования различных типов данных и удобную совместную работу пользователей системы. QlikView в полном объеме представлена на всех мобильных платформах, что позволяет получать доступ к аналитике в любое время дня за пределами рабочего места. При работе с QlikView вопрос производительности возникает только при работе с очень большим набором данных, но существует много возможностей для ее увеличения. Как и другие представленные в отчете BI решения, QlikView поддерживает возможность автоматического обновления данных отчетов, обновления дашбордов и автоматическую рассылку отчетов на электронную почту пользователям системы в виде графических и pdf-файлов.

QlikView представляет собой модульную BI систему, состоящую из следующих компонентов: QlikView Desktop – для создание отчетов, QlikView Server и Publisher – для обмена отчетами и их публикации и клиентские приложения QlikView для различных платформ. В дополнение к существующим модулям для QlikView существует ряд расширений и дополнений, которые могут быть установлены дополнительно, исходя из потребностей пользователей системы: GeoQlik, 1C Коннектор, SAP Коннектор, R Коннектор, QVSource и другие.

Tableau BI

Tableau на сегодняшний день является одним из лидеров среди BI платформ. Из всех представленных на рынке BI решений Tableau преуспели больше всех в упрощении механизма доступа бизнес-пользователей к анализу данных.

Платформа была разработана в Стэнфордском университете в 2003 году. Разработчики Tableau провели множество исследований, чтобы построить систему, которая поддерживает естественную способность людей мыслить визуально [36]. В основе Tableau лежат технологии VizQL и Data Engine. Технология VizQL преобразует действия пользователей с графическим интерфейсом платформы в запросы к базе данных, которые обрабатываются и отображаются графически при помощи средств визуализации. Технология Data Engine позволила снять ограничение с хранилищ данных и ускорила их загрузку, что позволило строить интерактивные визуальные отчеты. Цель платформы Tableau, как визуализатора данных, — помочь пользователю понять его данные самыми эффективными способами из возможных.

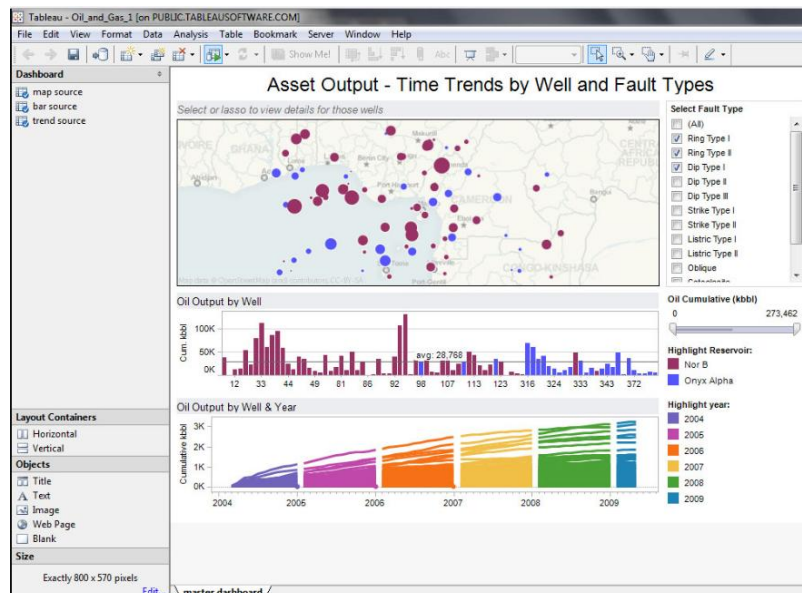


Рис.8. Рабочее окно Tableau

В состав продукта входит Tableau Desktop – модуль, в котором пользователь может конструировать отчеты, объединять их в информационные панели (дашборды) и в дальнейшем делиться ими с другими пользователями системы. Кроме этого, в состав продукта входит Tableau Server, который обеспечивает доступ всех пользователей системы через веб-браузер с их персональных компьютеров или мобильных устройств, а также позволяет встраивать отчеты в другие веб-приложения.

Tableau Desktop имеет интуитивно понятный пользовательский интерфейс, который позволяет автору отчетов быстро и эффективно работать с данными, объединять различные результаты анализа в дашборды, давая конечному пользователю возможность использовать фильтры и детализацию данных для быстрого понимания сути анализируемых процессов и явлений. Пользователи могут экспортировать данные в графические файлы и файлы pdf-формата, а также в Excel таблицы. Каждый отчет может быть индивидуально построен для отдельного пользователя, который может поделиться своими выводами и заключениями со своими коллегами.

Tableau обладает инструментами, которые позволяют на основе анализа данных строить прогнозы и тренды о развитии организации и предсказывать с определенной вероятностью различные события. Tableau предлагает отличную функциональность, которая включает постоянно обновляемые карты и дополнительную информацию для геопространственной аналитики. Tableau позволяет использовать данные из множества СУБД одновременно, а также поддерживает интеграцию с инструментами для работы с “Большими данными” (Big data). Tableau Desktop интегрирована с языком R – специальный язык программирования и среда разработки в открытом доступе для статистического анализа данных.

Prognoz Platform

Prognoz это российская компания, основанная в Перми в 1991 году на базе кафедры экономической кибернетики Пермского государственного

университета. Prognoz Platform имеет высокую популярность на территории России и Восточной Европы и используется как частными компаниями, так и органами государственной власти. Prognoz является единственной российской компанией, чье BI решение было представлено в рейтинге “Магический квадрант платформ бизнес-аналитики” компании Gartner, самом престижном в сфере BI.

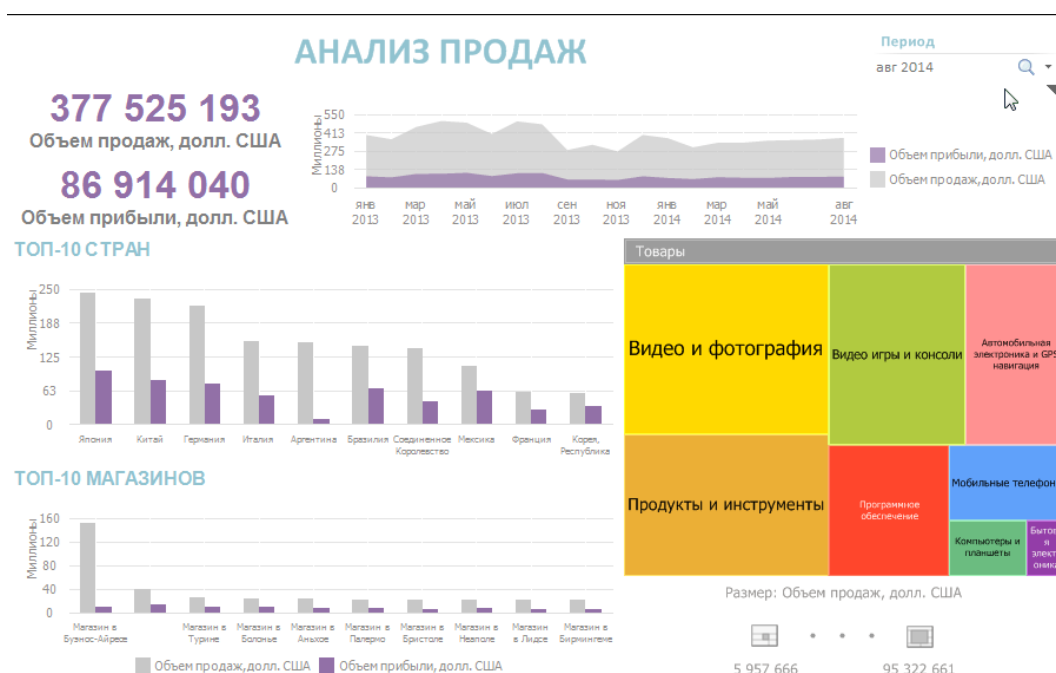


Рис.9. Рабочая область Prognoz Platform

Пользовательский интерфейс конструктора отчетов Prognoz Platform напоминает внешне продукты Office компании Microsoft, что обеспечивает интуитивную понятность в работе пользователя с инструментами платформы [34]. При этом платформа отлично интегрируется с приложениями Microsoft Office: результаты анализа можно экспортировать в Excel и Word, с хранилищами данных можно работать непосредственно из Excel, используя аналитические инструменты платформы. Пользователи имеют доступ к аналитическим инструментам Prognoz Platform через настольные и веб-

приложения, а также через мобильные устройства и облачную архитектуру системы.

Продукт Prognoz оснащен всеми стандартными инструментами для сбора и анализа данных, существующими на сегодня, а также продвинутыми инструментами моделирования и прогнозирования. С помощью Prognoz Platform можно получать, верифицировать и консолидировать большие объемы данных из разнородных источников. В Prognoz Platform реализованы самые последние достижения сферы бизнес-аналитики, включая технологии In-Memory (обработка данных в оперативной памяти устройства), Data Mining (интеллектуальный анализ данных), Search-Based BI (построение запросов в текстовом виде), а также Collaborative Decision Making (интегрированные инструменты совместного принятия решений).

Prognoz Platform поддерживает большинство современных средств визуализации данных, в том числе интерактивные 3D-карты. Через мобильное приложение для iOS также доступен широкий спектр функциональностей: поддержка динамического отображения данных в самых разных разрезах и видах, инструмент OLAP, средства анализа временных рядов. Prognoz Platform обладает интегрированной средой разработки, которая обеспечивает возможность быстрого создания персонализированных приложений.

Microsoft Power BI

Решение компании Microsoft в сфере BI состоит из трех высоко-интегрированных между собой составляющих: SQL Server, SharePoint Online и Power BI для Office 365. Вместе данные три продукта предоставляют полный спектр возможностей для просмотра, изучения, анализа и публикации данных [32].



Рис.10. Рабочая область Microsoft Power BI

Приложение Excel позволяет изучать и анализировать данные в интерактивных представлениях, таких как диаграммы и таблицы. С помощью Excel можно создавать интерактивные отчеты, системы показателей и панели мониторинга, а также предоставлять доступ к ним другим пользователям. Надстройки Power Query, Power Pivot, Power View и Power Map позволяют находить и объединять данные из различных источников, а также создавать многофункциональные и интерактивные представления и гибридные веб-приложения в Excel. С помощью служб Excel можно просматривать и обновлять отчеты, а также взаимодействовать с ними через браузер посредством предоставления доступа к централизованно управляемым отчетам, опубликованным на сайте SharePoint.

Power BI для Office 365 обеспечивает дополнительные возможности бизнес-аналитики в облаке:

Настройка запланированного обновления данных для отчетов.

Просмотр и использование отчетов большого размера (до 250 МБ) на сайтах Power BI в Power BI для Office 365.

Использование запросов на естественном языке для поиска и анализа данных с помощью Power BI Q&A.

Предоставление доступа к запросам и источникам данных и управление ими.

Доступ к локальным данным из облака.

Возможность загрузки приложения Power BI из магазина Windows для просмотра книг на планшетах с Windows 8.

Tibco Spotfire

Spotfire – быстроразвивающаяся BI платформа от компании Tibco. Платформа Spotfire оперирует данными в оперативной памяти и имеет набор инструментов для интерактивной визуализации данных и расширенной аналитики [37]. Tibco Spotfire широко используется в различных сферах производства и науки для повышения уровня эффективности.

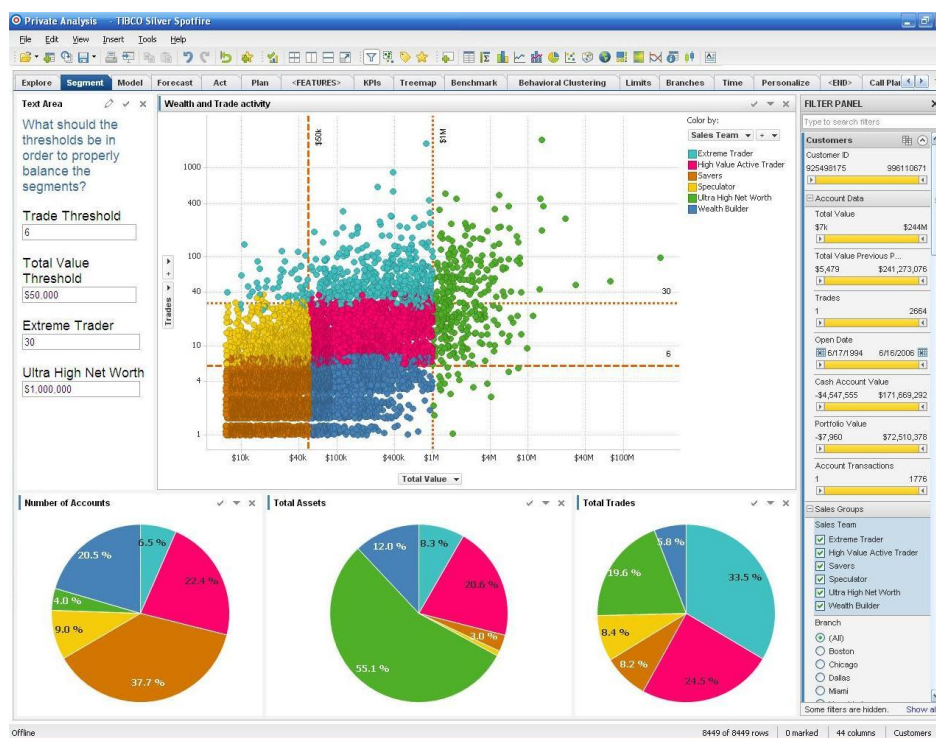


Рис.11. Вид рабочей области Tibco Spotfire

Платформа объединяет данные различных средств управления процессами и отображает их в удобном для руководителя виде. Платформа позволяет выявлять отклонения от требуемых показателей выполнения процессов и их причины, оценивать эффективность исполнителей. На основании анализа принимаются решения об оптимизации и оцениваются результаты оптимизации.

Spotfire поддерживает объединение данных из нескольких источников и отображение разработанных аналитических отчетов и моделей на основе этих данных через веб-браузер. Отчеты представлены в виде интерактивных дашбордов и аналитических приложений. В рамках отчета существует возможность перехода от статистического анализа нескольких сотен тысяч объектов к рассмотрению параметров конкретного объекта. Также для получения продвинутой аналитики возможна установка ряда специальных дополнений:

Spotfire®Statistical Services & TIBCO Enterprise Runtime for R – для прогнозирования и построения статистических моделей;

TIBCO GeoAnalytics – геопространственная аналитика в облачном сервисе;

Spotfire® Event Analytics – автоматизированный анализ различных данных с учетом исторического контекста;

Spotfire® Automation Services – сервис для автоматической рассылки аналитических отчетов.

IBM Cognos

IBM Cognos – это платформа для бизнес-аналитики, поставляемая одним из крупнейших мировых IT вендоров, компанией IBM [31]. Решение Cognos BI представлено целой линейкой программных продуктов для бизнес-аналитики: Enterprise, Insight, TM1, Express.

IBM Cognos BI предоставляет пользователям системы отчеты, аналитические данные, сводные информационные панели и карты

показателей, обеспечивая поддержку аналитических и рабочих процессов в организации. Cognos BI поддерживает функции совместного создания отчетов, которые позволяют бизнес-пользователям взаимодействовать с другими пользователями для принятия решений и получения дополнительной аналитики.



Рис.12. Рабочая область IBM Cognos

Пользователи могут получать доступ к отчетам на мобильных устройствах и работать с каждой страницей по мере загрузки отчета, а не ожидать загрузки полного отчета на мобильное устройство. Набор инструментов и данных на информационной панели может быть персонализирован для каждого пользователя для поддержки индивидуальных стилей принятия решения. С помощью функций карт показателей, имеющихся в Cognos BI, организации могут сопоставлять ключевые индикаторы производительности (KPI) со своими стратегическими целями.

Решения IBM Cognos не зависят от типа используемых СУБД, средств построения хранилищ данных, ERP-систем, операционных систем и браузеров. Решения IBM Cognos эффективно работают одновременно с фактически неограниченным количеством источников данных различных

типов (от крупнейших хранилищ данных и ERP-систем до текстовых файлов и файлов формата MS Office). Гибкий доступ к данным обеспечивает доступ к реляционным, OLAP - и локальным источникам данных в любой комбинации через использование централизованно управляемых моделей данных и метаданных. Мощный программный сервер оперативной аналитики позволяет оперативно создавать запросы и делать выкладки независимо от размера или сложности набора данных

IBM Cognos представляет собой набор интегрированных базовых функций, которые распределены по всем программным модулям. Например, в состав IBM Cognos Express входят три модуля, способные функционировать как в составе единого решения, так и самостоятельно:

- IBM Cognos Express Reporter для формирования отчетов и выполнения произвольных запросов;
- IBM Cognos Express Advisor для гибкого анализа и визуализации данных;
- IBM Cognos Express Xcelerator, обеспечивающий планирование и анализ бизнес-информации в Microsoft Excel.

Deductor Academic 5.2

Deductor Academic 5.2 – это платформа для анализа данных. Программа позволяет проводить кластеризацию, визуализировать и прогнозировать данные, применять различные подходы к их обработке и исследованию [24]. Бесплатный вариант платформы включает в себя программу анализа Studio в комплекте с многомерной базой данных Warehouse и предназначается для использования в исключительно образовательных, некоммерческих целях.

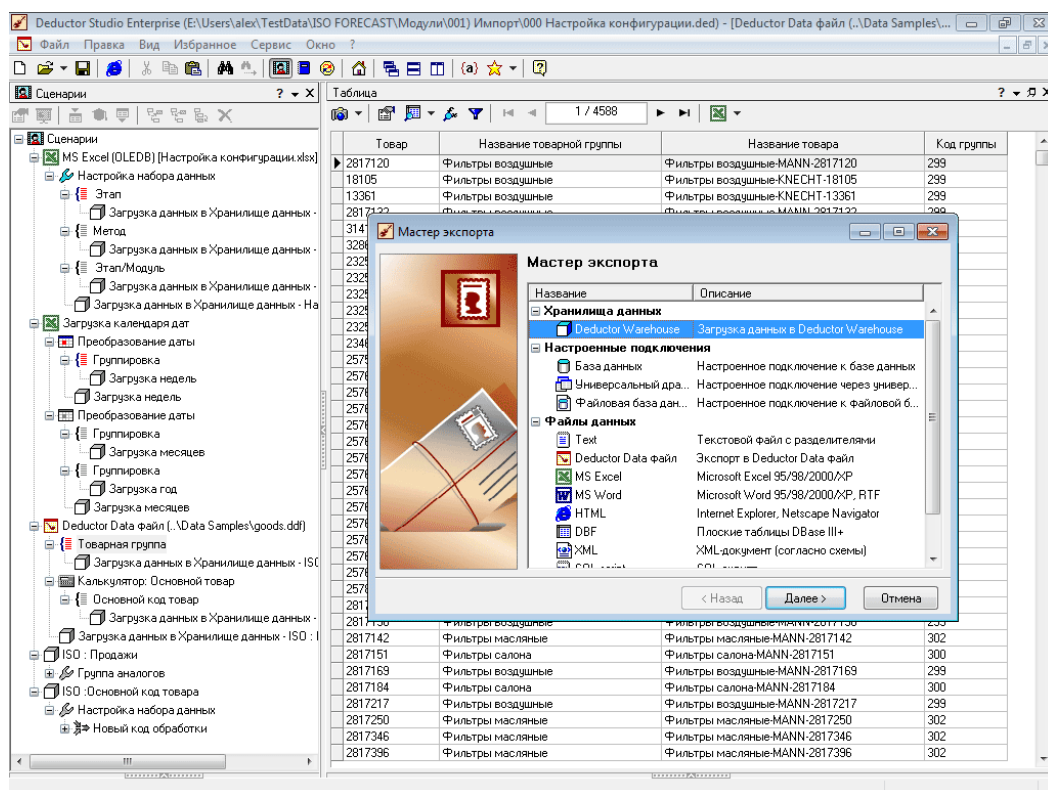


Рис.13. Рабочая область Deductor Academic

Программа поддерживает экспорт и импорт данных из текстовых файлов, позволяет строить различные модели (в т. ч. – на основе нейронных сетей, линейной регрессии и др.) и визуализировать полученные результаты в виде таблиц, диаграмм, графиков и т. д. Deductor содержит все необходимое для решения задач Data Mining. Применение методов Data Mining – фактически единственная возможность извлечь пользу из накопленной информации, в противном случае собранные данные будут лежать «мертвым грузом». Data Mining позволяет извлекать из данных знания и превратить в конкурентные преимущества: качественно прогнозировать, точнее, выявлять целевые аудитории, предсказывать развитие событий, управлять рисками.

В результате изучения выбранных VI-платформ была получена следующая сравнительная таблица.

Таблица 1. Сравнение BI платформ

Критерий сравнения	Qlik View	Tableau	Spotfire	Prognoz	MS Power BI	Deductor Academic	IBM Cognos
Некоторые сравнительные характеристики							
Наличие официального сайта на русском языке	V	-	-	V	V	V	-
Наличие бесплатной демо-версии	V	V	V	V	V	V	V
Наличие русифицированных курсов	V	V	-	V	V	V	V
Оценки в результате тестовой эксплуатации, по 5-ти бальной шкале							
Оценка техподдержки	5	5	5	5	0	5	4
Оценка масштабируемости	3	4+	5	4	4	4	4
Оценка клиентского доступа	5	5	5	4	2	5	5
Оценка интерфейса	4	4+	4	4	3	4	2
Оценка интегрируемости	4	5	4	3	3	4	4
Оценка визуализации	4 -	4	4	3	4	5	3
Оценка администрирования	4	5	4	4	4	4	3
Оценка поддержки OLAP	2	5	5	5	4+	5	5
Итоговые оценки по результатам предварительно изучения, по 5-ти бальной шкале							
Итоговая оценка перспектив внедрения	3	5	4	5	3	5	2
Итоговая оценка сложности системы	4	5	4	4	4	5	2

По результатам проведенного анализа наилучшую оценку получила программа Deductor Academic Studio 5.3. Связано это в первую очередь с тем, что эта программа является бесплатной, имеет понятный и удобный интерфейс и, не смотря на ограниченную функциональность бесплатной

версии, Deductor Academic имеет достаточно обширный набор методов предназначенных именно для интеллектуальной обработки данных.

Подводя итоги первой главы, можно выделить основные моменты:

Бизнес-анализ – достаточно молодое и быстро развивающееся направление в мире и на российском рынке. Для организаций любого масштаба будет полезным внедрение систем бизнес-аналитики. В настоящее время таких систем становится все больше, с каждым годом наблюдается прирост аналитических платформ, в том числе российского производства. При чем, большинством из них смогут воспользоваться пользователи разного уровня подготовки. Наиболее подходящей для организации ООО «Автолидер» была выбрана аналитическая платформа Deductor Studio Academic, так как имеет следующие привлекательные характеристики:

- наилучшим образом реализована визуализация данных;
- интуитивный пользовательский интерфейс;
- гибкость к частому изменению отчетности (разрезов, показателей);
- возможность обрабатывать и анализировать очень большие данные;
- российская разработка, которая имеет полноценный набор аналитических инструментов;
- бесплатная версия.

Глава 2. Информационные технологии для бизнес-анализа

2.1. Возможности аналитической платформы Deductor

Организации накапливают огромные массивы данных, однако не в состоянии получить от этой работы реальную отдачу. Имея исторические данные, можно решить критически важные для бизнеса задачи: оптимизировать процессы, управлять рисками, повышать доходность, удерживать клиентов...

Для таких задач не достаточно визуализации: отчетов, OLAP, информационных моделей - пользователи тонут в горах графиков. Для получения реальной отдачи нужно использовать методы глубокой аналитики (Data Mining), позволяющие не только просмотреть диаграммы, но автоматически находить закономерности, строить прогнозы, выявлять аномалии, т.е. помогать делать выводы.

Deductor - аналитическая платформа, в которой реализованы технологии позволяющие решить весь спектр задач полноценного анализа данных от консолидации и отчетности до прогнозирования и оптимизации.

Назначение

Решения на базе Deductor, позволяют пользователям быстро принимать обоснованные управленческие решения, благодаря реализации полного цикла аналитической обработки:

Консолидация данных из десятков разнородных источников

Очистка, систематизация и обогащение собранной информации

Отчетность, визуализация, OLAP-анализ, расчет KPI

Моделирование, прогнозирование, оптимизация

Самообучение на новых данных и адаптация моделей

Преимущества платформы Deductor

Для руководителя:

минимизация затрат: любая аналитика - одна платформа;

эволюционное развитие от простой отчетности до сложной оптимизации;

апробированная платформа - сотни проектов со сложной аналитикой;

десятки партнеров, тысячи обученных специалистов.

Для аналитика:

полный спектр технологий анализа: Data Warehouse, ETL, OLAP, Data Mining, KDD;

аналитика от простых формул до самообучающихся алгоритмов;

модификация логики анализа без привлечения программистов;

поддержка: система дистанционного обучения, кейсы, отраслевые курсы.

Для IT-специалиста:

встроенная интеграция с десятком источников данных;

высокая производительность;

техническая поддержка непосредственно от разработчиков.

Организационная структура аналитической платформы Deductor

Аналитическая платформа Deductor состоит из пяти частей:

1. Warehouse - хранилище данных, консолидирующее информацию из разных источников;
2. Studio - приложение, позволяющее пройти все этапы построения прикладного решения, рабочее место аналитика;
3. Viewer - рабочее место конечного пользователя, одно из средств тиражирования знаний (т. е. когда построенные аналитиком модели используют пользователи, не владеющие технологиями анализа данных);
4. Server - служба, обеспечивающая удаленную аналитическую обработку данных;

5. Client - клиент доступа к Deductor Server. Обеспечивает доступ к серверу из сторонних приложений и управление его работой. Разработчиками представляются три типа аналитической платформы Deductor: Enterprise; Professional; Academic. В нашем случае представлен пакет Academic. Версия Academic предназначена для образовательных и обучающих целей. Ее функционал аналогичен версии Professional за исключением:

- отсутствия пакетного запуска сценариев, работа в программе может вестись только в интерактивном режиме;
- отсутствия импорта из промышленных источников данных: 1С, СУБД, файлы MS Excel, Deductor Data File;
- также недоступны некоторые другие возможности.

Версия Enterprise предназначена для корпоративного использования и имеет расширенный потенциал.

Версия Professional предназначена для небольших компаний и однопользовательской работы. В ней отсутствуют серверные компоненты, поддержка OLE, виртуальное хранилище, а традиционное хранилище данных можно создавать только на СУБД FireBird. Автоматизация выполнения сценариев обработки данных осуществляется только через пакетный режим.

В представленной версии ввод информации в программу может быть осуществлен только путем импорта данных, сохраненных в текстовом формате. Сценарии не осуществляются в пакетном режиме, а все расчеты необходимо выполнять самостоятельно, запуская каждый сценарий. Разработчики платформы выделяют четыре уровня пользователей:

- аналитик;
- пользователь;
- администратор;
- программист.

Студенты могут выступать в качестве аналитиков и пользователей аналитической платформы. Функции аналитика:

- создание в Deductor Studio сценариев - последовательности шагов, которую необходимо провести для получения нужного результата;
- построение, оценка и интерпретация моделей;
- настройка панели отчетов для пользователей Deductor Viewer;
- настройка сценария на поточную обработку новых данных.

Функция пользователя - это просмотр готовых отчетов в Deductor Viewer.

Состав и назначение модулей

Deductor Warehouse - многомерное хранилище данных, предназначенное для решения задачи консолидации информации. Использование единого хранилища позволяет обеспечить простой и прозрачный доступ к данным, контроль целостности и непротиворечивости информации, высокую скорость обработки. Благодаря глубокой степени интеграции любую информацию из хранилища данных можно получить в приложениях Deductor с минимальными усилиями. Хранилище данных ориентировано именно на аналитическую обработку, поэтому включает в себя все, что необходимо для комфортной работы при анализе. Оно содержит интегрированный семантический слой, то есть механизм, автоматически преобразовывающий бизнес-термины в операции с базой данных и обратно. Благодаря наличию семантического слоя пользователь оперирует такими бизнес-понятиями, как «клиент», «товар», «прибыль», а система автоматически выполняет необходимые действия с базой данных и предоставляет пользователю нужную информацию. Применение хранилища данных позволяет не быть привязанным к учетной системе, хранить данные не только за последний период, а за весь необходимый для анализа срок, консолидировать информацию из разнородных источников. Использование специализированных методов хранения и извлечения данных значительно

увеличивает скорость получения информации. Хотя наличие единого источника данных не является обязательным условием работы аналитической системы, практически всегда ее создание начинается с построения хранилища данных.

Deductor Warehouse поддерживает прозрачную работу с тремя СУБД: Firebird, MS SQL и Oracle. Вне зависимости от используемой СУБД работа с хранилищем происходит совершенно одинаково с использованием единого унифицированного механизма доступа. Поддержка нескольких СУБД в качестве платформы хранилищ, данных позволяет в каждом конкретном случае применять наиболее пригодную для данного случая базу данных. В нашем случае используется бесплатное (FireBird) программное обеспечение. Кроме того, в Deductor реализована поддержка концепции виртуальных хранилищ, данных - Virtual Warehouse. Виртуальное хранилище данных обеспечивает прозрачный для аналитика доступ к сведениям, хранящимся в любых реляционных СУБД. Взаимодействие с Virtual Warehouse происходит аналогично работе с традиционным хранилищем данных. Аналитик оперирует бизнес-понятиями, заданными в семантическом слое, и от него скрыты все сложности выборки данных, как и в случае с Deductor Warehouse. Пользователь задает при помощи простого Мастера, какая информация его интересует, а система автоматически трансформирует их в запросы к базе данных. Таким образом, эмулируется работа хранилища данных, а данные реально не перегружаются в специализированную систему, все операции производятся «на лету». Virtual Warehouse позволяет представить информацию, хранящуюся в реляционных базах данных, в удобном для аналитика многомерном виде. Deductor Studio - это рабочее место аналитика. В этом приложении осуществляется формализация знаний эксперта. Программа включает все необходимые для анализа инструменты обработки: механизмы импорта данных из разнородных источников, методы очистки и

предобработки, алгоритмы построения моделей и механизмы экспорта данных.

Все действия по анализу данных сводятся всего к 4 операциям:

1. Импорт данных.

В процессе импорта данные получаются из источника и загружаются специальным образом в программу. В дальнейшем с ними можно производить любые доступные операции, работа со всякой импортированной таблицей происходит одинаково. Поддерживается импорт из наиболее распространенных СУБД (Oracle, MS SQL, MySQL, Interbase...), стандартных файлов обмена данными (dbf, txt, csv...), офисных приложений (MS Excel, MS Access...), бизнес-программ (1С v7, v8...). Кроме того, в программу встроен механизм импорта с применением стандартов доступа к данным ODBC и ADO.

2. Обработка данных.

Обработкой называется любое действие над данными, приводящее к их преобразованию, например, очистка данных либо построение моделей. Ее результатом является набор данных, который можно опять обработать каким-либо способом. Благодаря этому обеспечивается возможность построения сценариев обработки, то есть последовательных операций над данными, приводящих к нужному результату. Поддерживается широкий набор механизмов обработки: методы очистки (заполнение пропусков, редактирование аномалий, фильтрация...), инструменты предобработки (квантование, группировки, сортировки...), методы построения моделей (нейронные сети, самоорганизующиеся карты, деревья решений...).

3. Визуализация.

Полученные результаты можно просмотреть различными способами, начиная от простых таблиц и диаграмм до многомерных кубов и специализированных визуализаторов. Система построена таким образом, что

самостоятельно определяет возможные способы визуализации и предлагает наиболее удобные способы отображения данных для каждого случая.

4. Экспорт данных.

Результаты обработки могут быть выгружены во множество приемников данных. Таким образом, обработанная и проанализированная информация выходит за пределы аналитической платформы, попадает в бизнес-приложения, офисные программы и прочее. В Studio реализованы самые современные самообучающиеся алгоритмы анализа. Анализ данных в Deductor Studio базируется на построении сценариев обработки.

Алгоритм типового сценария представляется следующим образом:

1. аналитик загружает анализируемые данные в Excel или, в нашем случае, в текстовый редактор;
2. производит операцию импорта;
3. проверяет данные на наличие ошибок и исправляет их, например, продажи с нулевой суммой или возврат товара поставщику, этот процесс называется операцией очистки;
4. группируются данные для получения итоговой информации по месячным продажам определенного товара - это операции трансформации;
5. аналитик подбирает модель полинома или другую формулу, которые объясняли, исторические продажи - это этап построения модели;
6. применяется построенная модель для получения прогноза на следующий период, реализуется процесс прогнозирования;
7. последний этап анализа - отправка результатов прогноза заинтересованному лицу. Этот процесс реализуется экспортом полученных результатов. Работая с Deductor, аналитик строит сценарий по описанному алгоритму.

Deductor Studio не имеет механизмов ввода и ручной правки данных. В случае, если аналитик, получив данные, обнаружит в них, например, ошибки, он должен будет описать правило работы с такими данными. Например, он

должен будет отфильтровать данные о продажах с нулевой суммой. Это является обязательной операцией, так как вручную в Deductor Studio удалить непригодные записи невозможно. То, что он сформулирует, автоматически станет частью сценария. Такая работа требует чуть больше усилий и времени, чем простое удаление данных из таблицы, но подобный сценарий обработки тиражируем. При появлении новых данных не нужно опять искать некорректные записи, а воспользоваться правилом их обработки в сценарии, и очистка данных может быть выполнена автоматически. Эта особенность позволяет говорить о Deductor как об инструменте тиражирования знаний.

В Deductor сценарии отображаются в виде дерева с иконками и пояснительным текстом. Взглянув на это дерево, можно без труда проследить логику сценария и понять особенности его реализации. Это помогает не только модифицировать сценарии, но и передавать их другому аналитику, который также просто сможет «прочитать» ход мысли аналитика, создавшего сценарий.

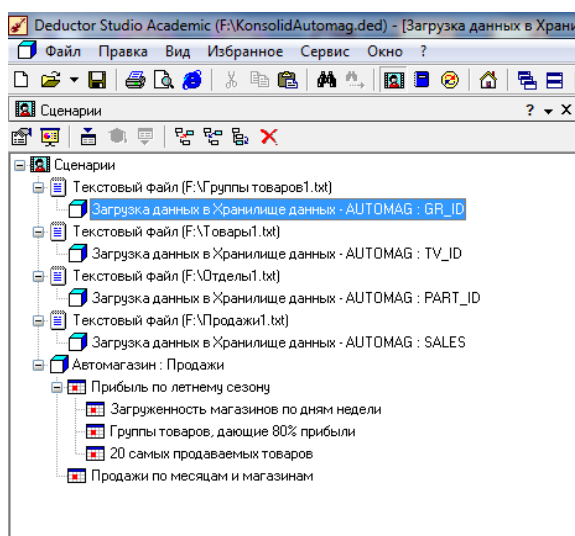


Рис.14. Дерево сценариев Deductor Studio

Анализ не ограничивается только обработкой данных, визуализация данных позволяет значительно повысить результативность анализа. В системе имеется множество удобных способов отображения данных. Программа самостоятельно анализирует способы обработки, особенности

набора данных, на которых производился анализ и автоматически предлагает возможные способы визуализации. Среди множества механизмов визуализации, встроенных в Deductor Studio, имеется и мощный Online Analytical Processing (OLAP) модуль. OLAP - один из наиболее популярных способов отображения табличных данных. Данные в этом случае могут отображаться в виде кросс-таблиц или кросс-диаграмм. Кросс-таблицы удобны тем, что большая часть операций манипулирования данными выполняется «на лету». Одним щелчком мыши, можно данные сгруппировать произвольным образом, отфильтровать, отсортировать, переставить столбцы/строки и произвести множество других операций. Deductor Studio позволяет при помощи этого механизма визуализации просмотреть любые данные, т. е. не только саму исходную информацию, но и результаты любой обработки.

Deductor Studio - это инструмент аналитика, а он является ключевым лицом в процессе анализа данных, именно его знания формализуются и тиражируются, но многие пользователи не являются аналитиками, для них нужен более простой и понятный способ получения требуемой информации. В Deductor Studio имеется панель отчетов, напоминающая проводник в известных операционных системах. На этой панели аналитик формирует иерархическую структуру папок и в определенные папки выносит ссылки на интересующие пользователей узлы сценария.

Deductor Viewer - это рабочее место конечного пользователя. В нем отсутствуют механизмы построения сценариев, настройки источников, данных и прочие сложности. Работа с программой упрощена до предела: пользователь видит настроенную аналитиком панель отчетов, выбирает интересующий отчет, программа автоматически выполняет все необходимые действия, и конечный пользователь получает результат. Эта составляющая является частью более расширенных видов аналитической платформы и в нашем случае не применяется, впрочем, как и Deductor Server/Client.

Deductor Server функционирует в виде Windows-службы, к которой можно обращаться удаленно при помощи специального клиента - Deductor Client. Управлять выполнением сценарием можно как из локальной сети, так и через Интернет. Использование Deductor Server значительно упрощает создание полноценной корпоративной аналитической системы, его применение позволяет воспользоваться всеми преимуществами трехзвенной архитектуры, оптимально используя возможности серверной аналитической обработки.

2.2. Бизнес-анализ деятельности торговой компании в аналитической платформе Deductor

Проектирование хранилища данных

Хранилище данных Deductor Warehouse – это специально организованная база данных, ориентированная на решение задач анализа данных и поддержки принятия решений, способная обеспечить максимально быстрый и удобный доступ к информации.

Все данные в Deductor Warehouse хранятся в структурах типа «снежинка», где в центре расположены таблицы фактов, а «лучами» являются измерения, причем каждое измерение может ссылаться на другое измерение.

На этапе проектирования структуры хранилища, потребуются данные, которые представлены в четырех таблицах (табл. 2-5).

Таблица 2. Группы товаров

Товарная группа.Код	Товарная группа
11	Автозапчасти
12	Автохимия
13	Автокосметика
14	Масла
15	Шины/Диски
16	Аккумуляторы
17	Автосвет
18	Автоаксессуары
19	Автоэлектроника
20	Инструмент
21	Сувениры

Таблица 3. Отделы

Отдел.Код	Отдел.Наименование
1	Колесо 1
2	Колесо 2
3	Колесо 3

Таблица 4. Товары (фрагмент)

Товар.Код	Товар.Наименование	Товарная группа
11001	Турбокомпрессор Hyundai Porter (НОВЫЙ) T0903	11
11002	Глушитель, задняя часть, BOSAL, 190875	11
11003	Глушитель, средняя часть, BOSAL, 177247	11
11004	Глушитель, задняя часть, BOSAL, 233563	11
11005	Глушитель, задняя часть, BOSAL, 154799	11
11006	Ремкомплект GPM, OPTIMAL, SK1031	11
11007	Ремкомплект GPM, DAYCO, KTB304	11
11008	Ремкомплект GPM, BOSCH, 1987949190	11
11009	Ремкомплект GPM, CONTITECH, CT605K1	11
11010	Свеча зажигания, DENSO, T14PRU15	11

Таблица 5. Продажи (фрагмент)

Дата продажи	Товар.Код	Отдел.Код	Час покупки	Количество	Сумма
01.06.2017	130207	1	10	1	175.50
01.06.2017	130209	1	12	2	251.60
01.06.2017	130208	1	12	1	171.00
01.06.2017	130504	1	15	1	99.80
01.06.2017	319236	1	17	2	540.30
01.06.2017	319237	1	17	3	790.50
01.06.2017	125111	1	18	2	362.90

Создание хранилища

Запускаем Deductor Studio Academic. Для создания нового хранилища данных – в меню Вид выбираем вкладку «Подключения», нажав правой кнопкой мыши, запускаем «Мастер подключений (Рис.15).

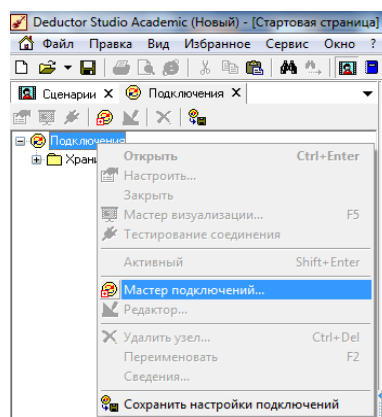


Рис. 15. Создание хранилища данных

Выбираем тип источника – Deductor Warehouse (Рис. 16).

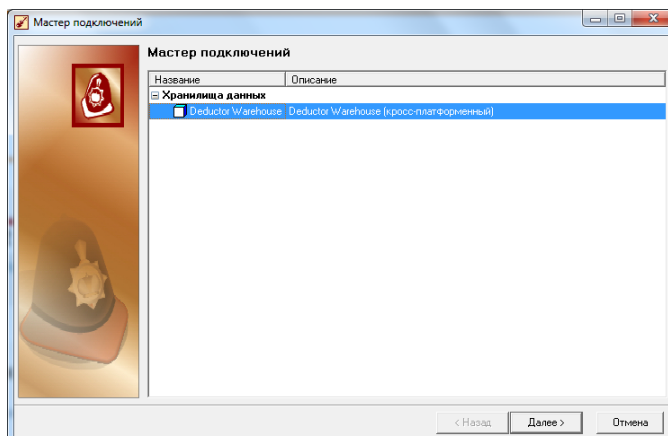


Рис. 16. Окно выбора типа источника подключения Deductor Warehouse

Следующим шагом – выбираем доступный тип базы данных – Firebird (В платных версиях платформы Deductor Studio есть возможность выбрать MS SQL или Oracle). Задаем параметры базы данных (далее БД), в которой будет создана физическая и логическая структура хранилища данных (Рис. 17). Для этого пропишем путь, где будет располагаться БД, устанавливаем логин/пароль, определяем нужную кодировку, поставим флажок – Сохранять пароль.

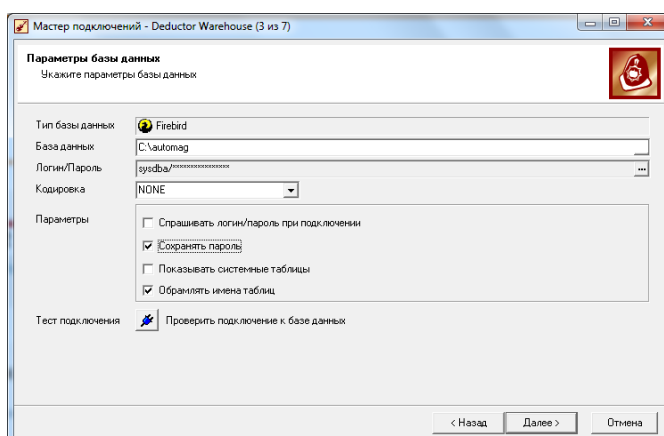



Рис. 17. Установка параметров базы данных automag.gdb

Затем выбираем необходимую версию Хранилища Данных для корректной работы с метаданными – Deductor Warehouse 6. Создаем файл БД с необходимой структурой метаданных, с помощью кнопки , в результате по указанному в предыдущем шаге пути будет создан файл automag.gdb. Имя хранилища вводится латинскими буквами, метку (название ХД видимое пользователю) установим «Автоматазин», в окне «Описание» - дадим описание ХД. Хранилище данных успешно завершено (Рис. 18).

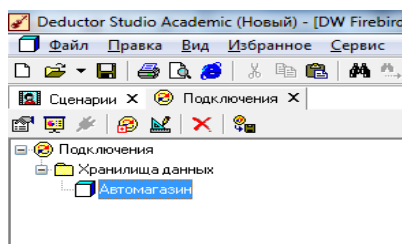



Рис. 18. Хранилище данных «Автомагазин»

Для отражения созданной структуры хранилища данных в хранилище, вызываем редактор метаданных , расположенный на вкладке «Подключения».

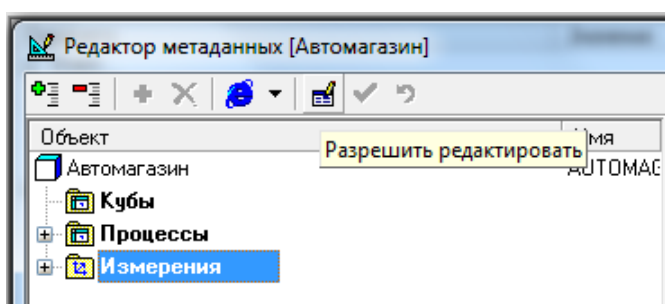


Рис. 19. Редактор метаданных [Автомагазин]

Внесем первое измерение «Код группы», для этого нажмем кнопку «Разрешить редактировать» (Рис. 19), выберем узел «Измерения» и добавим измерение Код группы (Рис. 20).

С помощью кнопки «Добавить» создаем измерения с необходимыми параметрами, приведенными в табл. 6. Таким образом, структура метаданных хранилища будет состоять из пяти измерений.

Таблица 6. Параметры измерений

Измерение	Имя	Метка	Тип данных
Код группы	GR_ID	Группа.Код	Целый
Код товара	TV_ID	Товар.Код	Целый
Код отдела	PART_ID	Отдел.Код	Целый
Дата	S_DATE	Дата	Дата/время
Час покупки	S_HOUR	Час	Целый

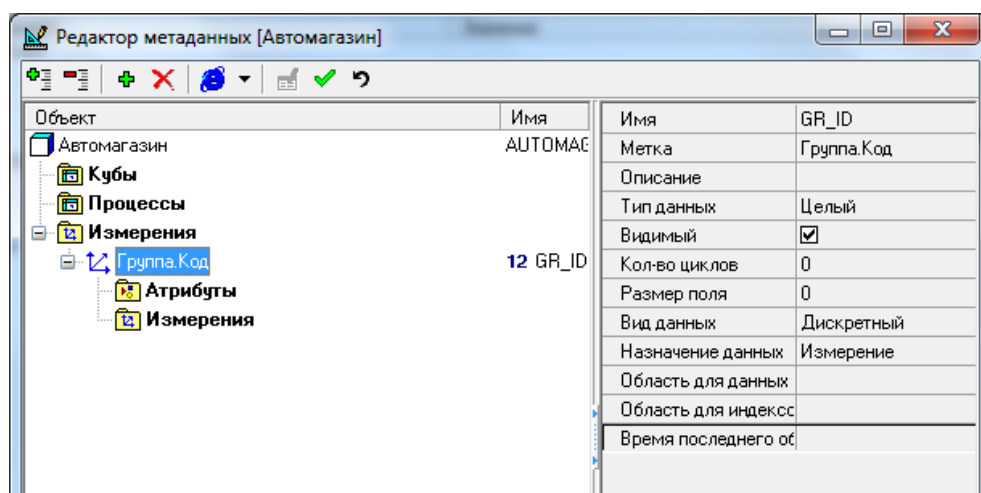


Рис. 20. Создание первого измерения «Код группы»

Каждое из измерений может ссылаться на другое измерение, тем самым реализуя иерархию измерений. Создадим ссылку измерения Товар.Код на измерение Группа.Код (Рис. 21).

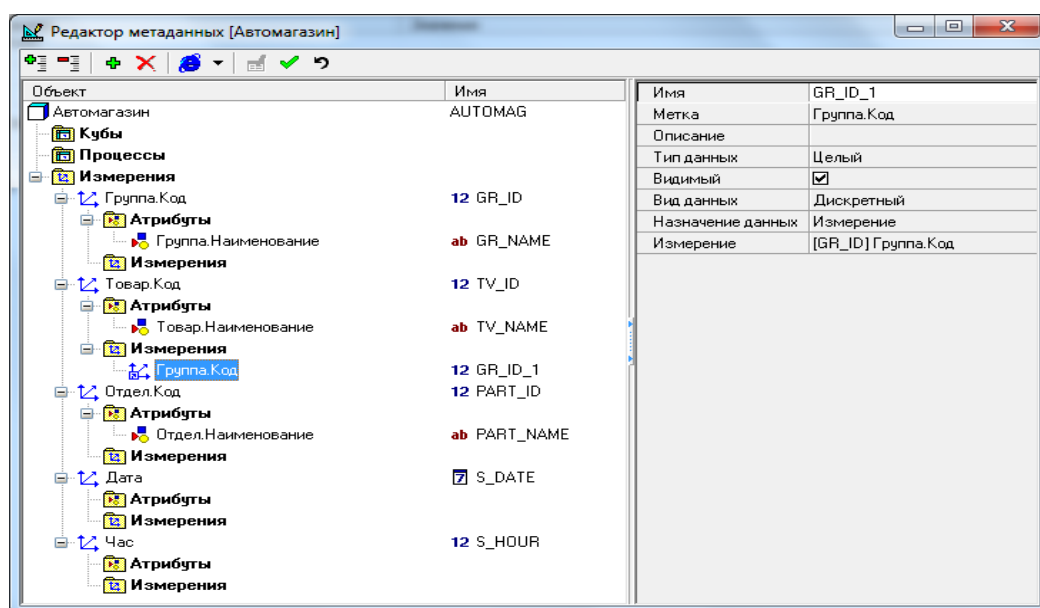


Рис. 21. Формирование ссылки на измерение «Группа.Код»

После создания всех необходимых измерений и ссылок на измерения, сформируем процесс «Продажи». Добавляем в этот процесс измерения: Дата, Отдел.Код, Товар.Код, Час и два факта: Количество и Сумма (Рис. 22).

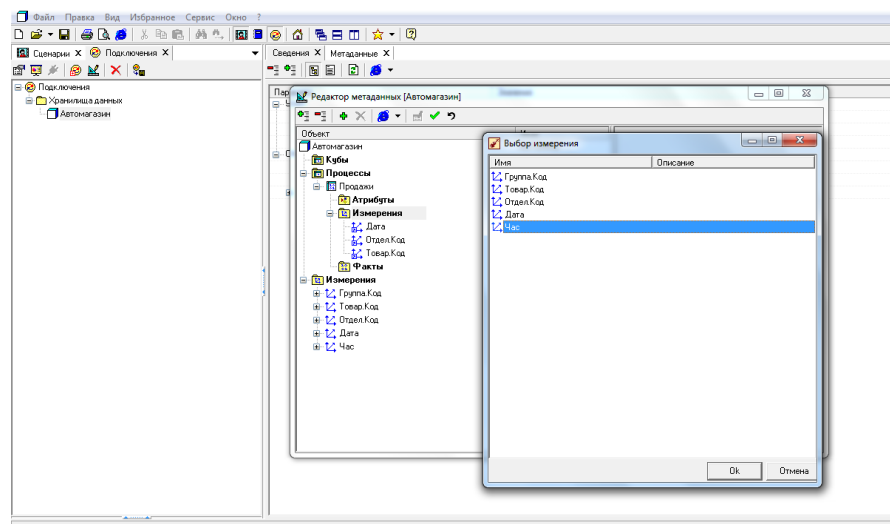


Рис. 22. Формирование процесса «Продажи»

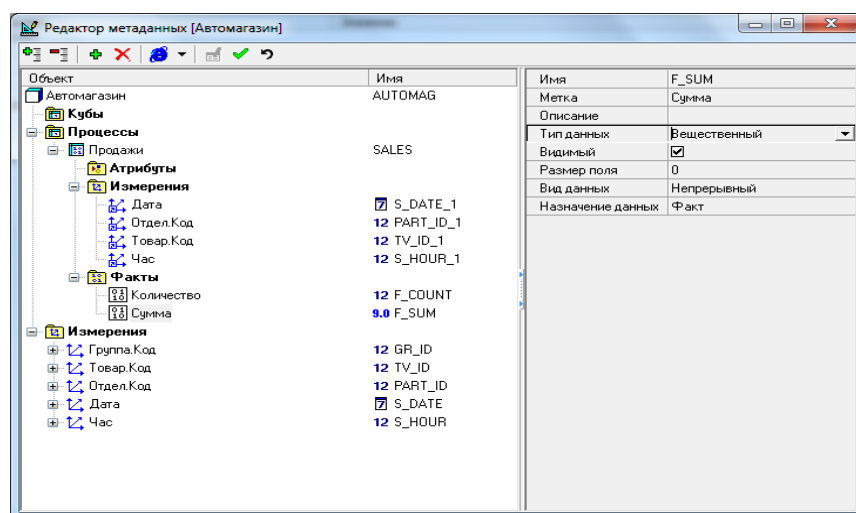



Рис. 23. Структура хранилища данных «Автомагазин»

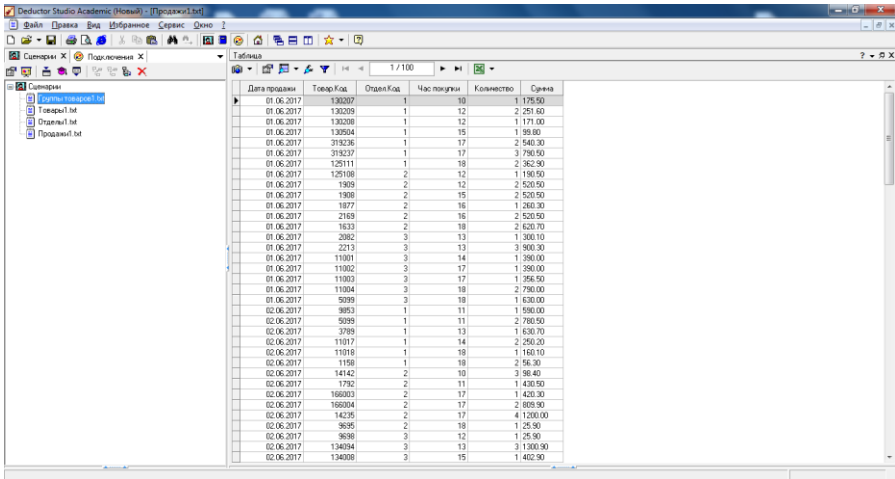
Получили необходимую структуру хранилища данных «Автомагазин» (Рис.23).

В результате всех вышеописанных действий получаем пустое ХД с настроенным семантическим слоем. Теперь необходимо загрузить в хранилище данные из внешних источников. Для этого создаем сценарий.

Важно знать, что при создании сценария сначала загружаются все измерения, имеющие атрибуты и только потом загружаются данные в процессы, измерения загружаются с верхнего уровня к нижнему.


Первым шагом сценария – будет Импорт данных. Так как в бесплатной версии Deductor Studio Academic возможно импортировать исключительно текстовые файлы, были подготовлены четыре файла с расширением txt: Группы товаров1.txt, Товары1.txt, Отделы1.txt, Продажи1.txt.

Для импорта файла во вкладке «Сценарии», нажимаем кнопку вызова «Мастер импорта»  - выберем источник – нужный текстовый файл и настроив параметры импорта, импортируем все четыре текстовых файла (Рис. 24).



Дата продажи	Товар.Код	Отдел.Код	Час.покупки	Количество	Цена
01.06.2017	130207	1	18	1	1 175.00
01.06.2017	130209	1	12	2	2 251.60
01.06.2017	130208	1	12	1	1 171.00
01.06.2017	130504	1	16	1	1 99.00
01.06.2017	319236	1	17	2	2 540.30
01.06.2017	319237	1	17	3	3 780.90
01.06.2017	125111	1	18	2	2 362.40
01.06.2017	125108	2	12	1	1 180.00
01.06.2017	1909	2	12	2	2 520.90
01.06.2017	1908	2	15	2	2 520.90
01.06.2017	1877	2	16	1	1 260.30
01.06.2017	2189	2	16	2	2 520.90
01.06.2017	1633	2	18	2	2 620.70
01.06.2017	2082	3	13	1	1 300.10
01.06.2017	2213	3	13	3	3 900.30
01.06.2017	11001	3	14	1	1 360.00
01.06.2017	11002	3	17	1	1 380.00
01.06.2017	11003	3	17	1	1 350.90
01.06.2017	11004	3	18	2	2 760.00
01.06.2017	5099	3	18	1	1 630.00
02.06.2017	9853	1	11	1	1 590.00
02.06.2017	5099	1	11	2	2 760.90
02.06.2017	3789	1	13	1	1 630.70
02.06.2017	11017	1	14	2	2 250.20
02.06.2017	11018	1	18	1	1 160.10
02.06.2017	1198	1	18	2	2 56.30
02.06.2017	14142	2	10	3	3 96.40
02.06.2017	1792	2	11	1	1 430.90
02.06.2017	166003	2	17	1	1 420.30
02.06.2017	166004	2	17	2	2 809.90
02.06.2017	14235	2	17	4	4 1200.00
02.06.2017	9695	2	18	1	1 25.90
02.06.2017	9696	3	12	1	1 25.90
02.06.2017	134094	3	13	3	3 1300.90
02.06.2017	134008	3	15	1	1 402.90

Рис. 24. Результат импорта текстового файла «Группы товаров1.txt»

Вторым шагом загрузим данные в измерения, с помощью «Мастера экспорта» . Из списка выбираем загрузка данных в Deductor Warehouse, затем выбираем необходимое подключение к хранилищу данных – «Автоматизация». Далее указываем измерение, в которое будет загружаться информация – *Группа.Код* (Рис. 25).

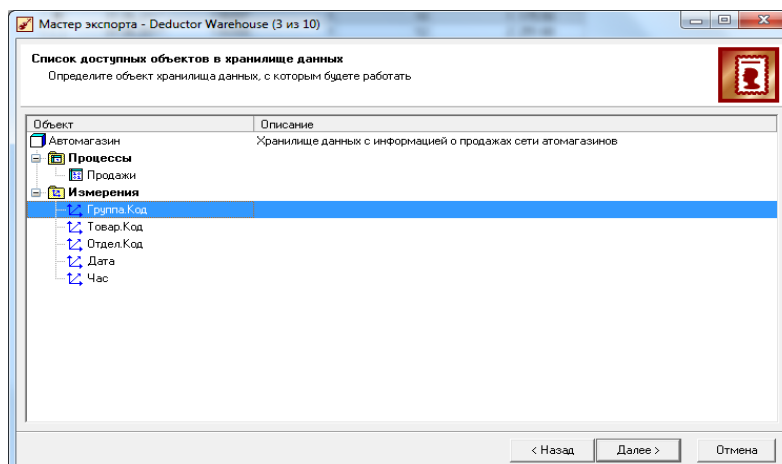


Рис. 25. Выбор объекта для экспорта

Устанавливаем соответствие элементов объекта в хранилище данных с полями входного источника данных (то есть таблицы Группы товаров1.txt (Рис. 26)).

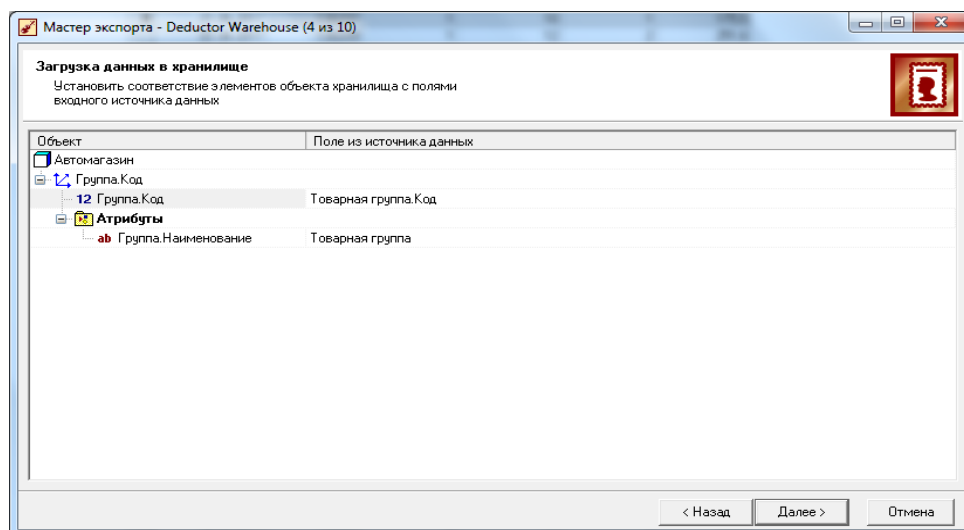


Рис. 26. Настройка соответствия полей

Запускаем процесс загрузки данных в ХД, нажав кнопку «Пуск». Повторим все действия для остальных измерений - *Отдел.Код*, *Товар.Код*, в результате получим сценарий, приведенный на рис. 27.

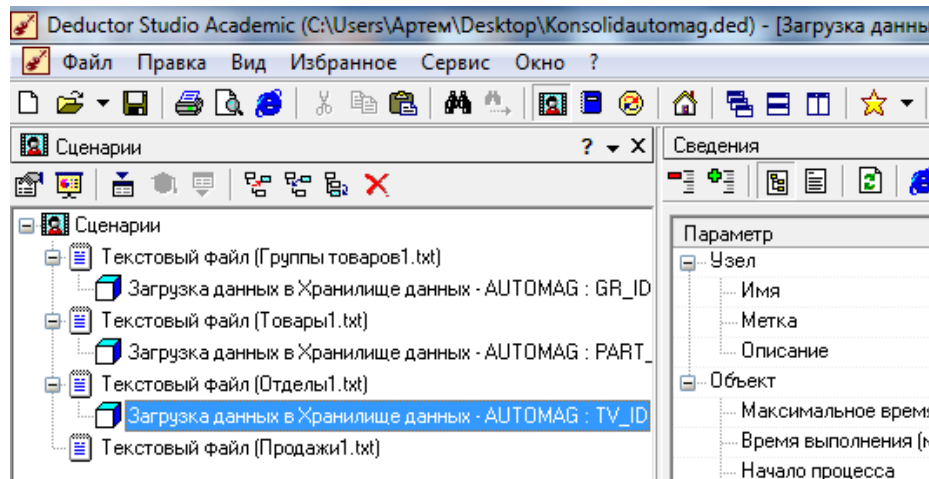


Рис. 27. Сценарий загрузки данных в ХД «Автомагазин»

Далее необходимо загрузить данные в процесс «Продажи» (Рис. 28).

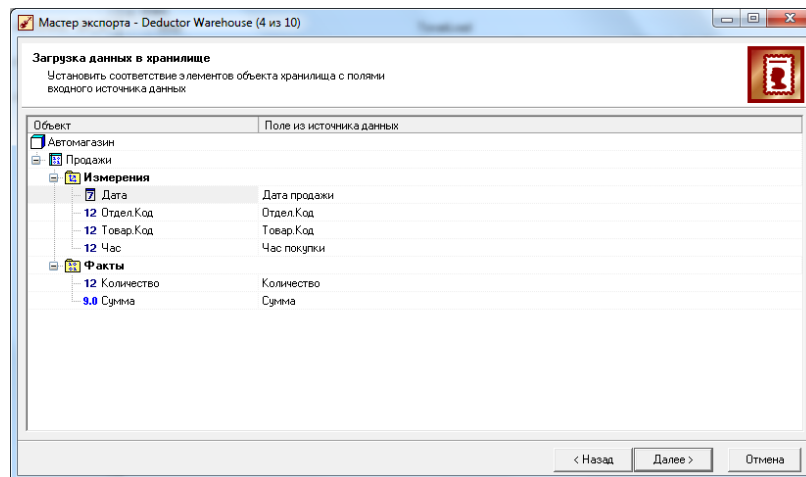


Рис. 28. Загрузка данных в хранилище

Окончательный результат загрузки:

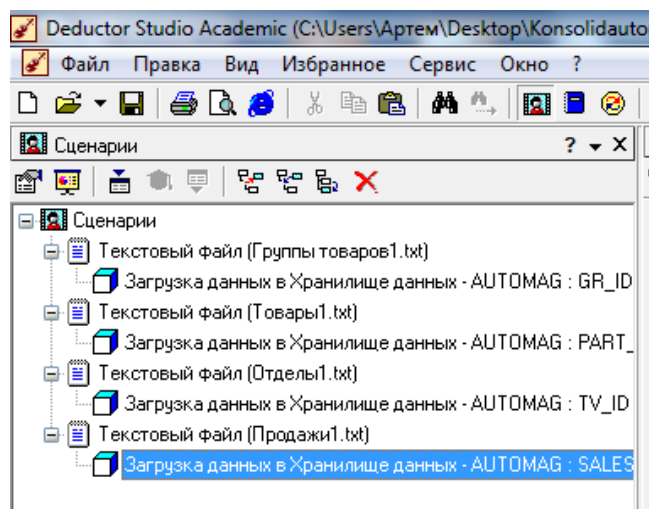


Рис. 29. Результат загрузки данных в хранилище

В результате проделанной работы получили: созданное и наполненное хранилище данных, и готовый сценарий загрузки (пополнения) информации из источников в ХД. Так же важно знать, что сценарий загрузки привязан не к самим данным, а к последовательности действий, которые необходимо осуществить для загрузки информации в хранилище. Таким образом, можно однажды создав сценарий, чтобы в будущем пополнять ХД.

Срезы из хранилища данных

С помощью Мастера импорта можно получить необходимые данные из ХД. Для того, чтобы построить отчет, который будет отражать изменения дохода от продаж по месяцам года в разрезе групп товаров и магазинов.

Вызываем Мастер импорта, выбираем тип источника данных – Deductor Warehouse, далее – хранилище «Автоматизация», а затем – процесс Продажи. Задаем необходимые для импорта измерения и атрибуты (Рис. 30). Благодаря иерархии внутри измерения Товар.Код появилась возможность доступа к измерению Группа.Код.

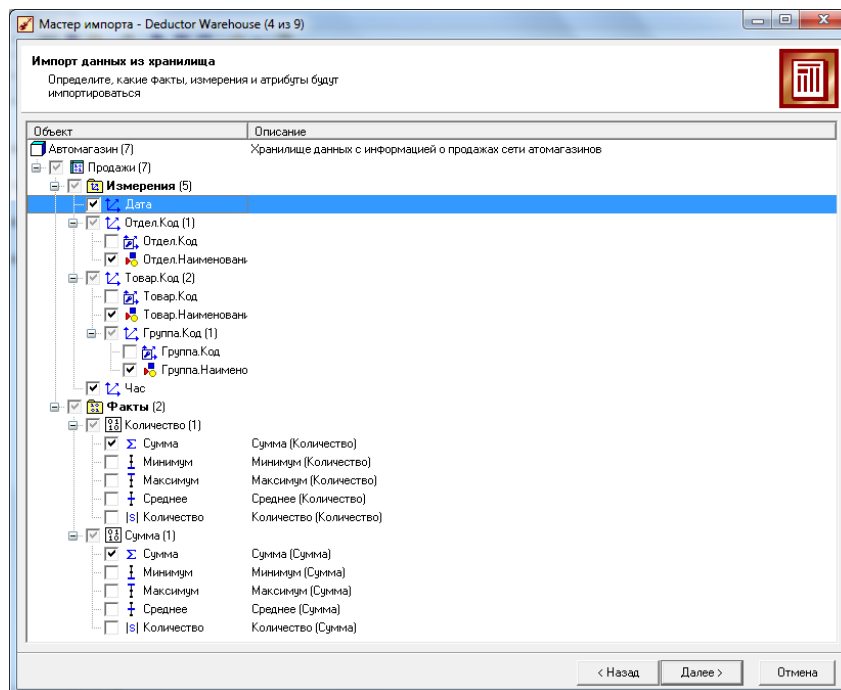


Рис. 30. Выбор импортируемых измерений, атрибутов и фактов

Определяем срезы для выбранных измерений. Это целесообразно делать при большом количестве значений измерения, так как позволяет загружать с сервера, на котором расположено ХД, только интересующие значения измерений и тем самым экономить время загрузки данных. Установим срез по измерению *Дата*: «Все продажи за последние 3 месяца от даты 09.08.2017» (Рис. 31-32).

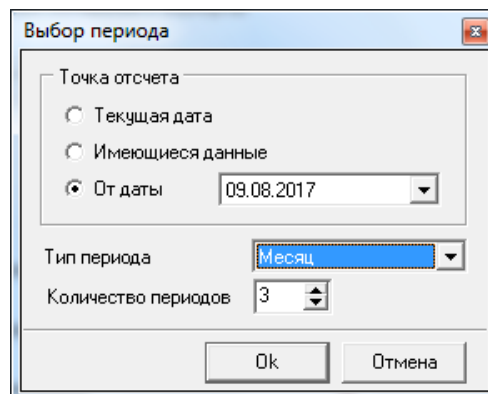


Рис. 31. Выбор периода

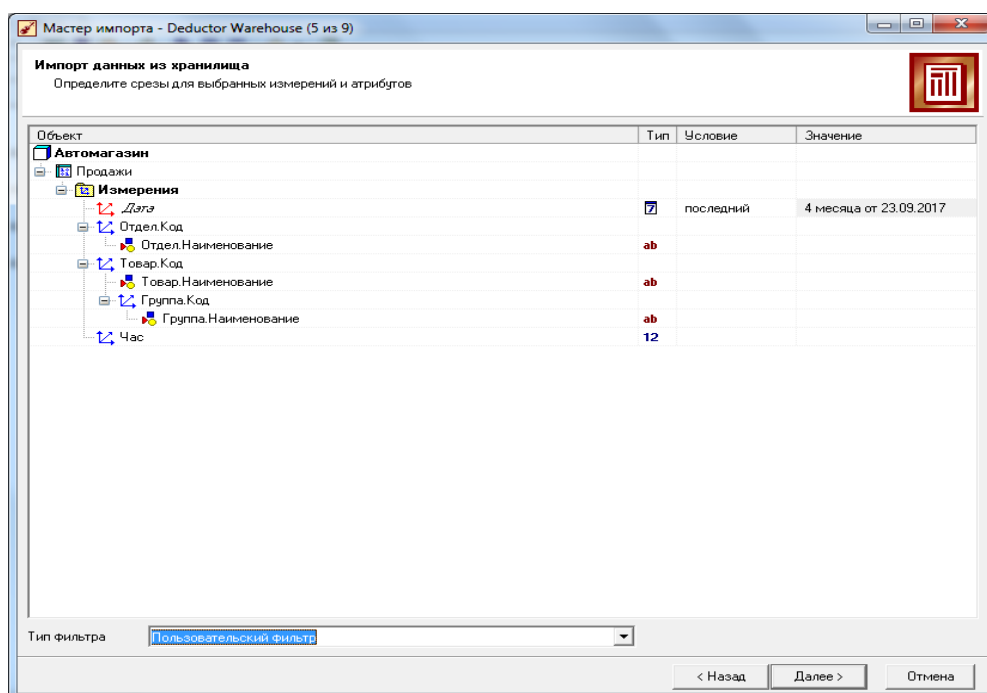



Рис. 32. Выбор срезов

Тип фильтра выставляем – Пользовательский фильтр. Это означает, что при каждом выполнении узла импорта будет выводиться окно, аналогичное окну настройки среза, в котором он сможет указать требуемые разрезы по этому измерению. Опция позволяет строить динамические отчеты, в которых пользователю предоставляется только интересующая его информация, а конкретные условия фильтрации он выбирает в момент импорта данных.

Нажав кнопку «Пуск», после импорта данных – выбираем визуализатор «Таблица».

Чтобы построить OLAP-отчет потребуются разделить измерение Дата на отдельные измерения *Месяц* и *Год*. Такую возможность дает обработчик «Дата и время» в Мастере обработки . Задача обработчика заключается в том, что на основе столбца с информацией о дате/времени формируются один или несколько столбцов, в которых указывается, к какому заданному интервалу времени принадлежит строка данных. Тип интервала задается на единственной вкладке настроек узла в зависимости от того, что необходимо выделить из даты (Рис. 33-34)

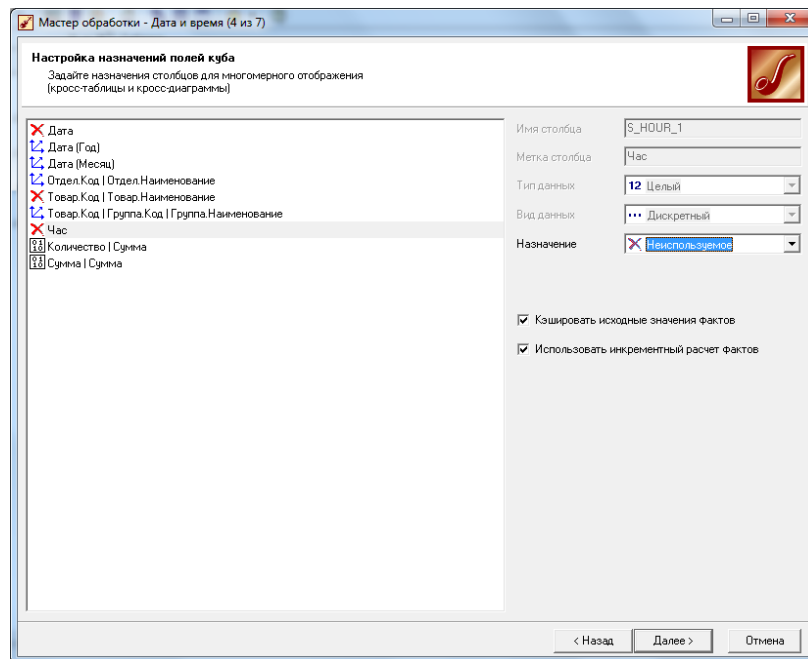


Рис. 33. Настройка назначений полей куба

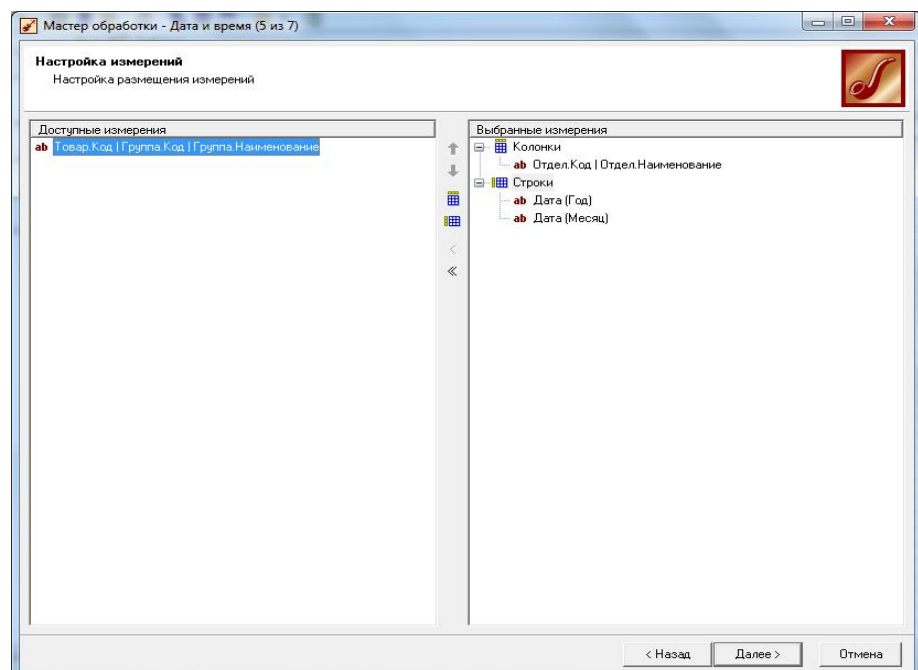


Рис. 34. Настройка размещений полей куба

В результате в выходном наборе будет создано два новых строчковых столбца с метками Дата (Год) и Дата (Месяц)(Рис. 35).

Определим группы товаров, которые приносят 80% выручки. Для этого в диалоговом окне «Селектор» (Рис. 37) выберем измерение Группа.Наименование, условие из выпадающего списка – Доля от общего, значение выставим – 80.

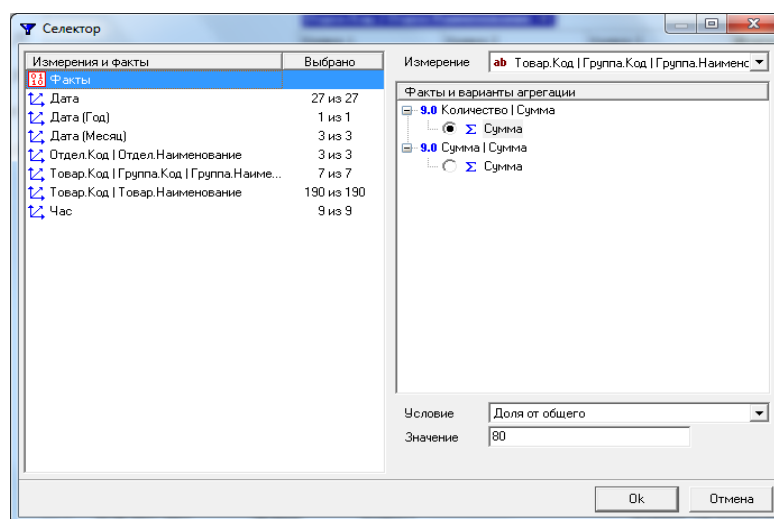


Рис. 37. Диалоговое окно «Селектор»

На рис. 38 изображен куб с заданными условиями и сортировкой суммы выручки по убыванию.

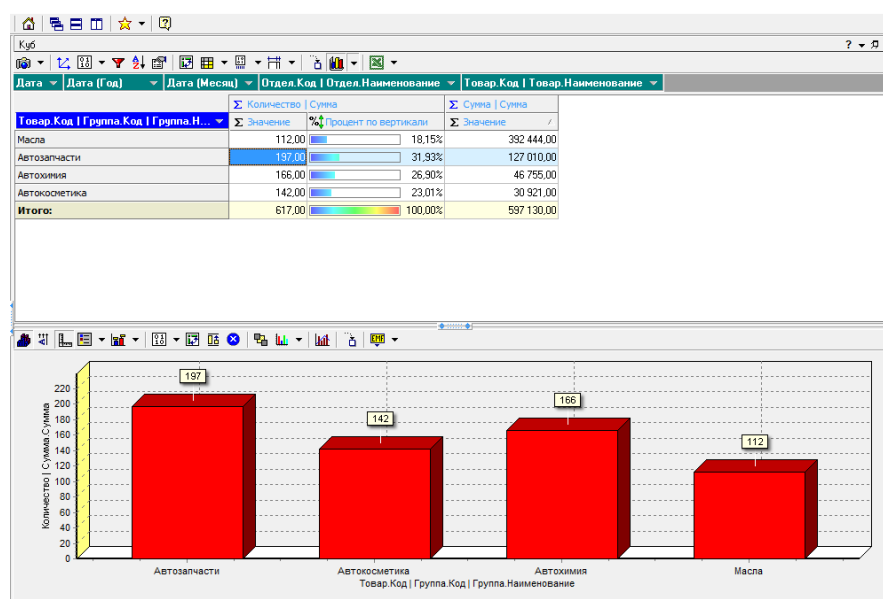


Рис. 38. Куб с заданными условиями и сортировкой суммы выручки по убыванию

Параллельно с кубом всегда строится кросс-диаграмма. Отличие кросс-диаграммы от обычной состоит в том, что она однозначно соответствует текущему состоянию куба и при любых его изменениях тоже меняется. На

рисунке 39 представленно кросс диаграмма по выручке трех магазинов за июнь, июль, август 2017 года.

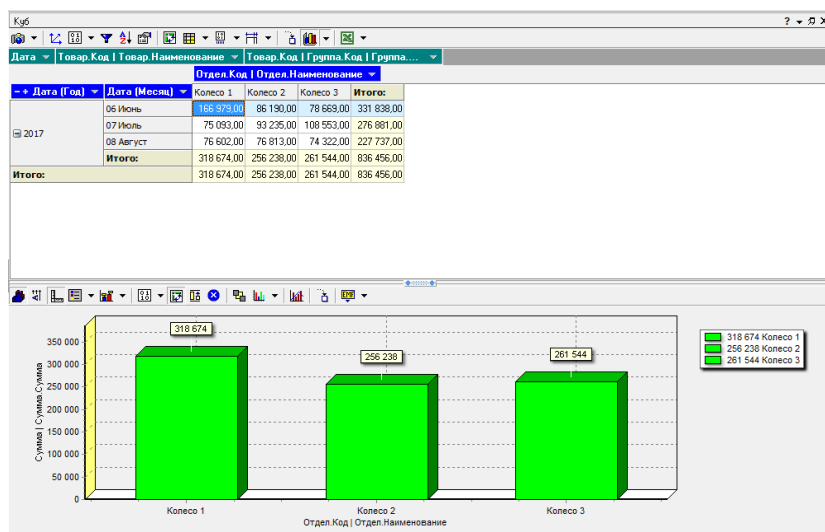


Рис. 39 кросс-диаграмма

Построим отчет и кросс-диаграмму загруженности магазинов (по количеству проданных товаров) за неделю (Рис. 40).

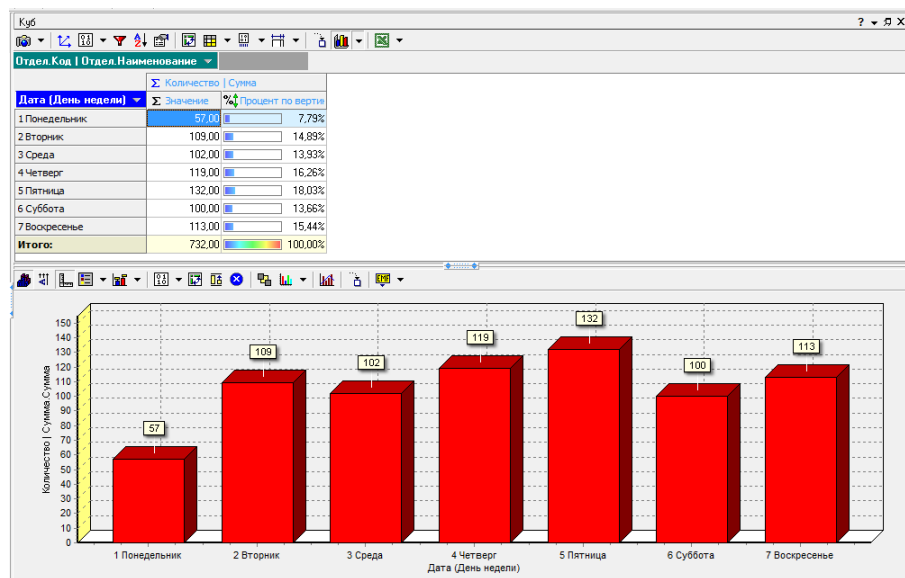


Рис. 40. Загруженность магазинов по дням недели

Таким образом, из наполненного хранилища данных, при помощи Мастера импорта появляется возможность строить необходимые многомерные отчеты, кросс-диаграммы. Осуществлять фильтрацию данных в кубе по заданным измерениям или фактам. Такая отчетность обеспечивает быстрый доступ к результатам и позволяет принимать правильные управленческие решения.

2.3. Ассоциативные правила в стимулировании розничных продаж

Выявление ассоциаций

В Deductor Studio для решения задач ассоциации используется обработчик Ассоциативные правила, в котором реализован алгоритм Apriori. Узел требует, чтобы на входе было два поля: идентификатор транзакции (номер чека, ID товара) и элемент транзакции (наименование товара), поля должны быть дискретного вида.

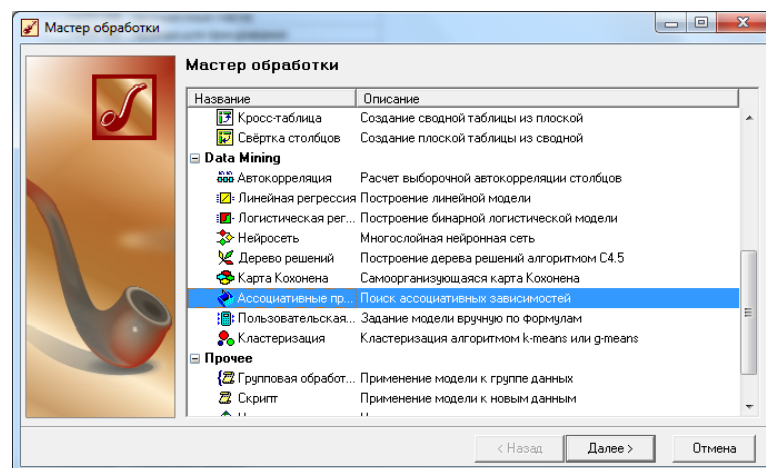


Рис. 41. Обработчик Ассоциативные правила

Создадим новый проект и импортируем данные из подготовленного текстового файла Чеки1.txt. В текстовом файле содержится таблица, состоящая из двух столбцов ID и ITEM (Рис.42).

Таблица	
ID	ITEM
S051184	масла моторные
S051184	воздушный фильтр
S051184	промывочные масла
S051184	провода для прикуривания
S051188	эмаль
S051200	грунтовка
S051200	провода для прикуривания
S051215	масла моторные
S051215	воздушный фильтр
S051216	масла моторные
S051216	маслоуловители
S051216	поддоны противоотливные
S051216	прокладки маслянного поддона
S051216	клапанные крышки
S051205	масла моторные
S051205	поддоны противоотливные
S051205	прокладки маслянного поддона

Рис. 42. Фрагмент таблицы из импортируемого файла Чеки1.txt.

Затем в Мастере обработки выберем пункт «Ассоциативные правила» (Рис. 42). Поле ID назначим идентификатором транзакции, а ITEM – ее элементом (Рис. 43).

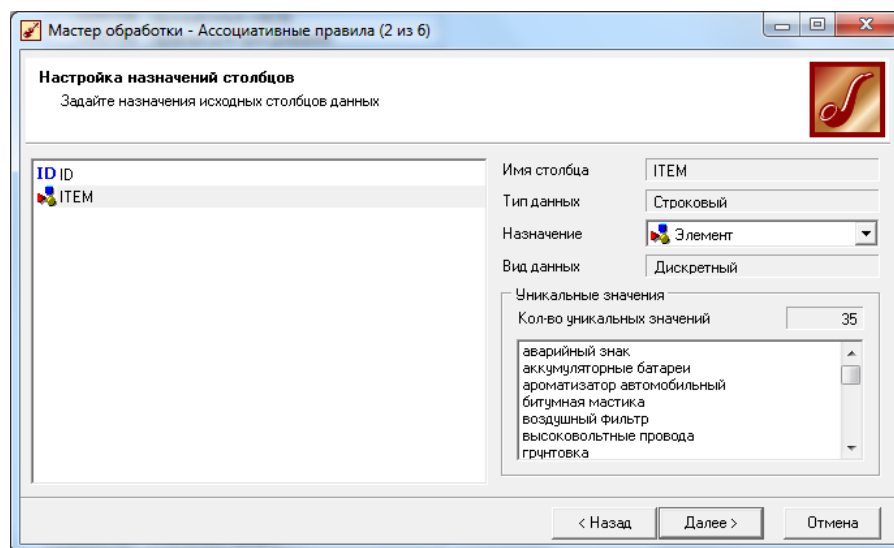


Рис. 43. Настройка назначения столбцов

В следующем окне необходимо настроить параметры построения ассоциативных правил (Рис. 44).

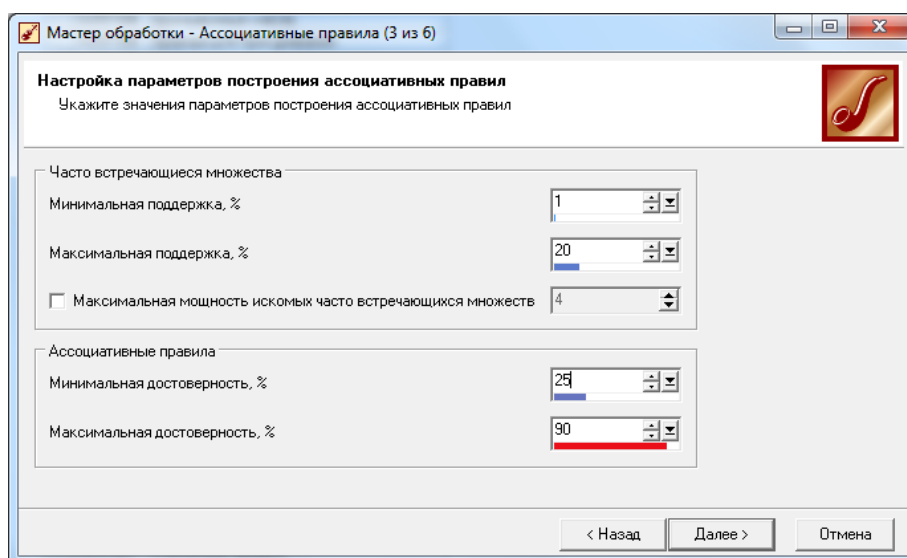


Рис. 44. Настройка параметров построения ассоциативных правил

1. Минимальная и максимальная поддержка, % – задают границы поиска часто встречающихся множеств, которые определяют множество популярных наборов, или частых предметных наборов, из этих множеств и создаются ассоциативные правила.

2. Минимальная и максимальная достоверность, % – в результирующий набор попадут только те ассоциативные правила, которые удовлетворяют условиям минимальной и максимальной достоверности.

3. Максимальная мощность искоемых часто встречающихся множеств – параметр ограничивает длину k-предметного набора. Например, при установке значения 4 шаг генерации популярных наборов будет остановлен после получения множества 4-предметных наборов. В конечном итоге это позволяет избежать появления длинных ассоциативных правил, которые трудно интерпретируются.

Далее нажимаем кнопку «Пуск», после чего программа запускает работу алгоритма поиска ассоциативных правил. В результате получим следующую информацию (Рис. 45):

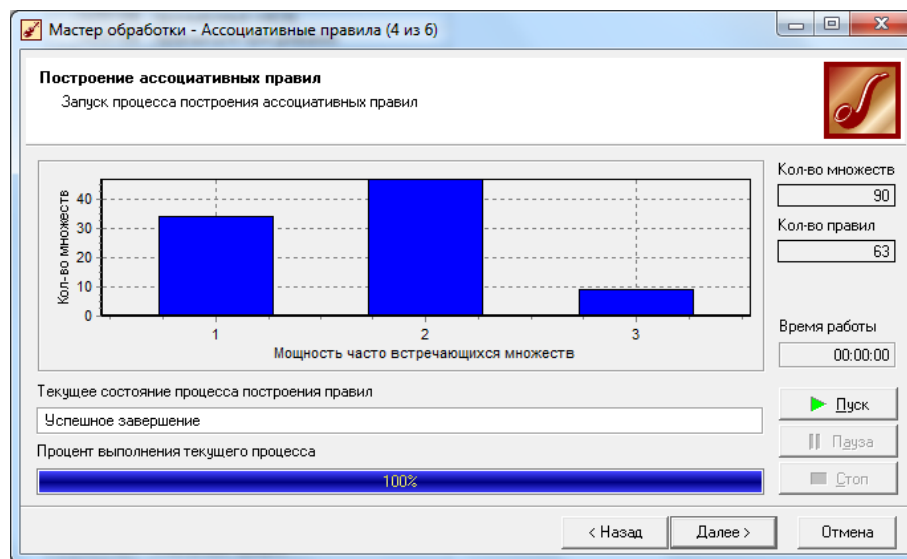


Рис. 45. Построение ассоциативных правил

Кол-во множеств – число популярных наборов, удовлетворяющих заданным условиям минимальной поддержки и достоверности (90 наборов);

Кол-во правил – число сгенерированных ассоциативных правил (найдено 63 правила).

В следующем шаге выбираем способы отображения данных: «Правила», «Популярные наборы», «Дерево правил» (Рис. 46).

Правила X Популярные наборы X Дерево правил X							
Правил: 63 из 63 Фильтр: Без фильтрации							
№	Номер правила	Условие	Следствие	Поддержка		Достоверность	Лифт /
				Кол-во	%		
1	43	воздушный фильтр	масла моторные	34	1,66	25,56	15,376
2	3	чехлы для сидений	ароматизатор автомобиль	79	3,86	54,86	12,064
3	2	ароматизатор автомобиль	чехлы для сидений	79	3,86	84,95	12,064
4	57	поддоны противоотливные	масла моторные	56	2,74	28,00	10,225
5	41	электросвет/лампы	провода автомобильные Г	28	1,37	31,82	7,311
6	40	провода автомобильные Г	электросвет/лампы	28	1,37	31,46	7,311
7	39	свечи зажигания	провода автомобильные Г	53	2,59	25,48	5,855
8	38	провода автомобильные Г	свечи зажигания	53	2,59	59,55	5,855
9	45	тормозные жидкости	свечи зажигания	53	2,59	58,89	5,790
10	44	свечи зажигания	тормозные жидкости	53	2,59	25,48	5,790
11	7	эмаль	битумная настилка	38	1,86	29,46	5,683
12	6	битумная настилка	эмаль	38	1,86	35,85	5,683
13	51	предохранители	воздушный фильтр	28	1,37	36,36	5,591
14	52	клапанные крышки	моторчик стеклоочистителя	21	1,03	38,89	5,232
15	11	промывочные масла	воздушный фильтр	96	4,69	32,88	5,055
16	10	воздушный фильтр	промывочные масла	96	4,69	72,18	5,055
17	49	воздушный фильтр	промывочные масла	28	1,37	71,79	5,028
18	46	воздушный фильтр	промывочные масла	34	1,66	70,83	4,961
19	24	промывочные масла	масляный фильтр	76	3,72	26,03	4,883

Рис. 46. Способы отображения данных

На вкладке «Популярные наборы» в виде списка отображается множество найденных популярных предметных наборов, которые можно отфильтровать и отсортировать. Зададим в фильтре минимальное значение поддержки 5% и отсортируем записи по ее убыванию, получим следующие 20 популярных наборов (Рис. 47).

Правила X Популярныe наборы X Дерево правил X						
Множеств: 20 из 90 Фильтр: Минимальная поддержка = 5,00						
№	Номер множества	Элементы	Поддержка		Мощность	
			Кол-во	%		
1	22	прокладки масляного поддона	392	19,17	1	
2	23	промывочные масла	292	14,28	1	
3	20	предохранители	288	14,08	1	
4	12	масла моторные	221	10,81	1	
5	26	свечи зажигания	208	10,17	1	
6	18	поддоны противоотливные	200	9,78	1	
7	9	клапанные крышки	199	9,73	1	
8	13	маслоуловители	195	9,54	1	
9	70	поддоны противоотливные	167	8,17	2	
10	24	радиатор охлаждения	154	7,53	1	
11	15	моторчик стеклоочистителя	152	7,43	1	
12	25	растворитель	146	7,14	1	
13	30	чехлы для сидений	144	7,04	1	
14	5	воздушный фильтр	133	6,50	1	
15	66	моторчик стеклоочистителя	131	6,41	2	
16	34	эмаль	129	6,31	1	
17	10	коврики для салона	119	5,82	1	
18	16	огнетушитель автомобильный	115	5,62	1	
19	14	масляный фильтр	109	5,33	1	
20	4	битумная настилка	106	5,18	1	

Рис. 47. 20 популярных наборов

На вкладке «Дерево правил» предлагается еще один удобный способ отображения множества ассоциативных правил. При построении дерева по

условию на первом (верхнем) уровне находятся узлы с условиями, а на втором – узлы со следствием. В дереве, построенном по следствию, наоборот, на первом уровне располагаются узлы со следствием.

Справа от дерева расположен список правил, построенный по выбранному узлу дерева (Рис. 48).

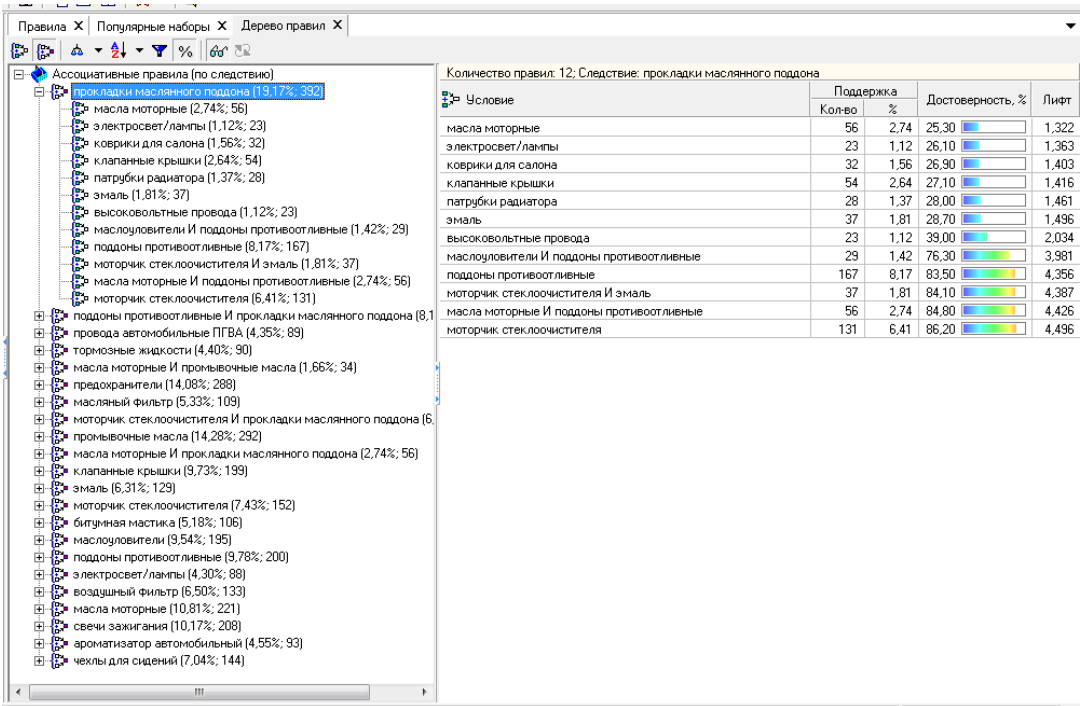


Рис. 48. Дерево ассоциативных правил

Для каждого правила отображаются поддержка и достоверность. Если дерево построено по условию, то вверху списка находится условие правила, а список состоит из его следствий. Тогда правила отвечают на вопрос: «Что будет если...».

Если же дерево построено по следствию, то вверху списка отображается следствие правила, а список состоит из его условий. Эти правила отвечают на вопросы: что нужно для того, чтобы получилось заданное следствие, или какие товары нужно продать для того, чтобы продать товар из следствия?

Правила X Популярныe наборы X Дeрево правил X					
Множествo: 20 из 30		Фильтр: Минимальная поддержка = 5,00			
№	Номер множества	ab. Элементы	Поддержка		Мощность
			Кол-во	%	
1	22	прокладки масляного поддона	392	19,17	1
2	23	промывочные масла	292	14,28	1
3	20	предохранители	288	14,08	1
4	12	масла моторные	221	10,81	1
5	26	свечи зажигания	208	10,17	1
6	18	поддоны противоотливные	200	9,78	1
7	9	клапанные крышки	199	9,73	1
8	13	маслоуловители	195	9,54	1
9	70	поддоны противоотливные	167	8,17	2
		прокладки масляного поддона			
10	24	радиатор охлаждения	154	7,53	1
11	15	моторчик стеклоочистителя	152	7,43	1
12	25	растворитель	146	7,14	1
13	30	чехлы для сидений	144	7,04	1
14	5	воздушный фильтр	133	6,50	1
15	66	моторчик стеклоочистителя	131	6,41	2
		прокладки масляного поддона			
16	34	зналь	129	6,31	1
17	10	коврики для салона	119	5,82	1
18	16	огнетушитель автомобильный	115	5,62	1
19	14	масляный фильтр	109	5,33	1
20	4	битумная мастика	106	5,18	1

Рис. 49. Часто встречающиеся множества с поддержкой более 5 %

Интерпретация ассоциативных правил

Самым важным этапом здесь является правильная интерпретация полученных правил, так как сами ассоциативные правила не являются результатом, готовым к использованию. Необходимо выделить правила, которые представляют интерес, определить какие из правил отражают закономерности и не являются ли артефактами. На этом этапе очень важно понимание предметной области.

Множества ассоциативных правил делятся на три вида.

Полезные правила содержат действительную информацию, которая ранее была неизвестна, но имеет логичное объяснение. Такие правила могут быть использованы для принятия решений, приносящих выгоду.

Тривиальные правила содержат действительную и легко объяснимую информацию, которая уже известна. Такие правила, хотя и объяснимы, но не могут принести какой-либо пользы, так как отражают или известные законы в исследуемой области, или результаты прошлой деятельности. При анализе рыночных корзин в правилах с самой высокой поддержкой и достоверностью

окажутся товары – лидеры продаж. Практическая ценность таких правил крайне низка.

Непонятные правила содержат информацию, которая не может быть объяснена. Такие правила получаются на основе или аномальных значений, или глубоко скрытых знаний. Напрямую эти правила нельзя использовать для принятия решений, так как их необъяснимость может привести к непредсказуемым результатам. Для лучшего понимания требуется дополнительный анализ.

Варьируя верхний и нижний пределы поддержки и достоверности, можно избавиться от очевидных и неинтересных закономерностей. Как следствие, правила, генерируемые алгоритмом, принимают приближенный к реальности вид. Значения верхнего и нижнего пределов сильно зависят от предметной области, поэтому не существует четкого алгоритма их выбора. Но есть ряд общих рекомендаций.

Большая величина максимальной поддержки означает, что алгоритм будет находить правила, хорошо известные или же настолько очевидные, что в них нет никакого смысла. Поэтому ставить порог максимальной поддержки очень высоким (более 20 %) не рекомендуется.

Большинство интересных правил находится именно при низком значении порога поддержки, хотя слишком низкое значение ведет к генерации статистически необоснованных правил. Поэтому правила, которые кажутся интересными, но имеют низкую поддержку, нужно дополнительно анализировать.

Уменьшение порога достоверности приводит к увеличению количества правил. Значение минимальной достоверности не должно быть слишком низким, так как ценность правила с достоверностью 5% чаще всего настолько мала, что это и правилом считать нельзя.

Правило с очень большой достоверностью ($> 85-90\%$) практической ценности в контексте решаемой задачи не имеет, так как товары, входящие в следствие, покупатель, скорее всего, уже приобрел.

Рассмотрим, правило № 6 масляный фильтр \rightarrow промывочные масла имеет $S = 3,72\%$; $C = 69,72\%$. Это говорит о том, что ожидаемая вероятность покупки набора промывочные масла \rightarrow масляный фильтр равна $3,72\%$. Если клиент приобрел масляный фильтр, то с вероятностью $69,72\%$ он выберет и промывочные масла.

Правила 17 из 17								
Фильтр: Без фильтрации								
№	Номер правила	Условие	Следствие	Поддержка		Достоверность	Лифт	
				Кол-во	%			
1	1	ароматизатор автомобиль	чехлы для сидений	79	3,86	84,95	12,064	
2	2	чехлы для сидений	ароматизатор автомобиль	79	3,86	54,86	12,064	
3	3	битумная мастика	свечи зажигания	50	2,44	47,17	4,638	
4	4	воздушный фильтр	промывочные масла	96	4,69	72,18	5,055	
5	5	комплект ГРМ	промывочные масла	42	2,05	48,28	3,381	
6	6	масляный фильтр	промывочные масла	76	3,72	69,72	4,883	
7	7	моторчик стеклоочистите.	прокладки масляного под.	131	6,41	86,18	4,496	
8	8	поддоны противоотливны	прокладки масляного под.	167	8,17	83,50	4,356	
9	9	прокладки масляного под.	поддоны противоотливны	167	8,17	42,60	4,356	
10	10	провода автомобильные Г	свечи зажигания	53	2,59	59,55	5,855	
11	11	тормозные жидкости	свечи зажигания	53	2,59	58,89	5,790	
12	12	воздушный фильтр масла моторные	промывочные масла	34	1,66	70,83	4,961	
13	13	воздушный фильтр предохранители	промывочные масла	28	1,37	71,79	5,028	
14	14	клапанные крышки прокладки масляного под.	поддоны противоотливны	25	1,22	46,30	4,734	
15	15	масла моторные поддоны противоотливны	прокладки масляного под.	56	2,74	84,85	4,426	
16	16	маслоуловители поддоны противоотливны	прокладки масляного под.	29	1,42	76,32	3,981	
17	17	моторчик стеклоочистите. эмаль	прокладки масляного под.	37	1,81	84,09	4,387	

Рис. 50. Правило № 6 масляный фильтр \rightarrow промывочные масла

С помощью анализа правил приходим к выводу, что многие правила предсказуемы, т.к. являются лидерами продаж сети магазинов, хотя в результате данного анализа можно выявить интересные правила. (например, свечи зажигания \rightarrow провода автомобильные). И тот факт, что при достоверности 54% встречаются тривиальные ассоциативные правила (например, чехлы для сидений \rightarrow ароматизатор автомобильный), говорит о том, что можно найти интересные правила при меньших значениях достоверности.

Для получения дополнительных неочевидных правил необходимо запустить алгоритм Apriori, устанавливаем интервал допустимой достоверности от 25 до 40%. В результате получим правила, которые можно назвать полезными: они неочевидны, но понятны. Например, возьмем правило провода автомобильные → электросвет/лампы.

В процессе внедрения и использовании аналитической платформы Deductor было спроектировано и разработано хранилище данных, которое впоследствии стало основой для внедрения системы аналитической OLAP-отчетности. Так же используя алгоритм Apriori, были извлечены ассоциативные правила, при помощи которых, можно предсказать какой из товаров будет покупаться, учитывая товар, который уже был приобретен ранее, так же определены товарные наборы, пользующиеся покупательским спросом.

На практике результаты анализа при помощи ассоциативных правил могут применяться в осуществлении кросс-продаж, учитываться при размещении товара на полках в магазинах, выявление наиболее выгодных товарных позиций и подбор наиболее продаваемых наборов товаров.

Заключение

В ходе дипломной работы мною были изучены основные концепции бизнес-анализа, описаны основные современные технологии, такие как Data mining, KDD, OLAP-технологии. Подробно описаны наиболее популярные на мировом рынке аналитические платформы, выбрана наиболее подходящая, с учетом специфики розничной торговли и, непосредственно, предприятия ООО «Автолидер», а именно система интеллектуального анализа данных – Deductor Academic.

В данной системе были разработаны шаблоны сценариев, которыми впоследствии компания может пользоваться на практике. С помощью этих сценариев организация имеет возможность консолидировать данные и делать OLAP-срезы о динамике продаж, загруженности магазинов, о самых покупаемых товарах, на основе спроектированного хранилища данных. Так же реализована возможность создавать необходимые для бизнеса отчеты.

При помощи ассоциативных правил выявлены часто выбираемые наборы товаров, выявлены наиболее выгодные товарные позиции.

В ходе дипломной работы были получены результаты, которые показывают, что при правильном выборе и внедрении системы интеллектуального анализа, бизнес получает существенные конкурентные преимущества и возможности для роста, посредством оптимизации бизнес-процессов.

Аналитическая платформа Deductor Studio содержит все необходимые функции, чтобы сделать бизнес-анализ прозрачным, управляемым и доступным.

Таким образом, платформа Deductor Studio доказывает свою жизнеспособность, эффективность и необходимость ее внедрения.

Список литературы

1. Абасова, С. Э. Информатизация профессиональной деятельности в сфере инженерного образования. Телекоммуникации и информатизация образования. [Текст]/ С.Э. Абасова. – СПб.: «Нестор». – 2013. – 428 с.
2. Аббакумов, В.Л. Бизнес-анализ информации. Статистические методы. [Текст]/ В.Л. Аббакумов, Т.А. Лезина. - М.: ЗАО «Издательство «Экономика». – 2012. – 374 с.
3. Абдикеев, Н.М. Управление знаниями корпорации и реинжиниринг бизнеса [Текст]/ Н.М. Абдикеев, А.Д. Киселев. – М.: Инфра-М. – 2013. – 384 с.
4. Бариленко, В.И. Основы бизнес-анализа [Текст]: учеб. пособие/ В.И. Бариленко, В.В. Бердников, Р.П. Булыга; под ред. В.И. Бариленко. – М.: Кнорус. – 2014. – 272 с.
5. Барсегян, А.А. Методы и модели анализа данных: OLAP и Data Mining. [Текст]: учеб. пособие / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И.Холод. – М.: ИНФРА-М, 2012. – 516 с.
6. Барсегян, А.А. Технологии анализа данных: Data Mining, Visual mining, Text mining, OLAP. [Текст]: учеб. пособие / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И.Холод. – СПб.: «БХВ Петербург», 2014. – 383 с.
7. Елиферов, В.Г. Бизнес-процессы: Регламентация и управление [Текст]: учеб. пособие / В.Г. Елиферов. – М.: НИЦ ИНФРА-М. – 2013. – 319 с.
8. Иванова, В.А. Путь аналитика. Практическое руководство IT-специалиста [Текст]/ В.А. Иванова, А.Д. Перерва. – СПб.: Питер. – 2016. – 304 с.
9. Ильина, О.П. Информационные системы и технологии в экономике и управлении [Текст]: учеб. для вузов/ О.П. Ильина, В.И. Кияев, А.П.

Приходченко, В.В. Трофимов; под ред. В.В. Трофимова. – М.: Издательство Юрайт. – 2015. – 542 с.

10. Матвеева, Т.А. Информационная культура [Текст] : учеб. пособие / Т.А. Матвеева, Г.А. Гейн. – Екб.: Центр «Учебная книга». – 2012.– 416с.

11. Паклин, Н.Б. Бизнес-аналитика: от данных к знаниям [Текст]: учеб. пособие/ Н.Б. Паклин, В.И. Орешков – СПб.: Питер. – 2013. – 704 с.

12. Рафалович, В. Data mining или интеллектуальный анализ для занятых [Текст]: практическое пособие/ В. Рафалович – М.:SmartBook. – 2014. – 96 с.

13. Репин, В.В. Бизнес-процессы. Моделирование, внедрение, управление [Текст] / В.В. Репин. – М.: Манн, Иванов и Фербер. – 2013. – 512 с.

14. Кукьер, К. Большие данные [Текст]: пер. с англ. Гайдюк И./ К. Кукьер, В. Майер-Шенбергер – М.: Издательство «Манн, Иванов и Фербер». – 2014. – 310 с.

15. Лесковец, Р. Анализ больших наборов данных [Текст]: пер. с англ. Слинкин А.А./ Р. Лесковец, А. Раджараман, Дж. Ульман – М.: ДМК Пресс. – 2016. – 498 с.

16. Фрэнкс, Б. Укрощение больших данных. Как извлекать знания из массивов информации с помощью глубокой аналитики [Текст]: пер. с англ./ сост. А.В. Баранов. – М.: Издательство «Манн, Иванов и Фербер». – 2014. – 352 с.

17. Артемьев В.И. Что такое Business Intelligence [Электронный ресурс]/ В.И. Артемьев// 2013. URL: <http://citforum.ru/consulting/Bi/whatis/>

18. Гудков Д.А. BI-платформы [Электронный ресурс]/ Д.А. Гудков// IT для бизнеса. 2014. URL: http://ko.com.ua/gartnerovskie_kvadraty_po_bi-platformam_2014_prodlzhenie_47749

19. Коваленко К.Г. Современные Business Intelligence (BI) системы на примере IBM Cognos BI [Электронный ресурс]/ К.Г. Коваленко// Интернет журнал Хабрахабр. 2015. URL: <https://habrahabr.ru/post/248829/>
20. Парамонов, С.П. Что такое Business Intelligence [Электронный ресурс]/ С.П. Парамонов// Блог компании Navicon. 2015. URL: <https://habrahabr.ru/company/navicon/blog/250875/>
21. Сорокина, Т.И. Бизнес-анализ предприятия: потребности компании [Электронный ресурс]/ Т.И. Сорокина// Журнал «Коммерческий директор». 2016. URL: <https://www.kom-dir.ru/article/1732-biznes-analiz>
22. Большие данные [Электронный ресурс]. – URL https://ru.wikipedia.org/wiki/Большие_данные
23. Основы бизнес-аналитики и науки о данных [Электронный ресурс]. – URL: <https://basegroup.ru/service/learning/catalog/bg101>
24. Сайт программы Deductor Academic 5.2 [Электронный ресурс]. URL: <http://www.basegroup.ru/deductor/>
25. Системы бизнес-аналитики (BI) [Электронный ресурс]. – <http://optima.ru/services/22/991/>
26. Системы бизнес-анализа (Business Intelligence) [Электронный ресурс]. – http://fogsoft.ru/solutions/business_intelligence/business_intelligence.aspx
27. BABOK Guide 2.0 [Электронный ресурс]/ URL: <https://www.iiba.org>
28. Business Intelligence. Системы бизнес-анализа: базовые понятия и решения [Электронный ресурс]/ URL: <http://1solution.ru/sobytiya/stati/sistemy-biznes-analiza-business-intelligence-bazovye-ponyatya-resheniya.html>
29. Business Intelligence [Электронный ресурс]/ URL: https://ru.wikipedia.org/wiki/Business_Intelligence

30. Data Mining [Электронный ресурс]/ URL:
https://ru.wikipedia.org/wiki/Data_mining
31. IBM Cognos [Электронный ресурс]/ URL:
<https://www.ibm.com/analytics/ru/ru/technology/products/cognos-analytics>
32. Microsoft Power BI [Электронный ресурс]/ URL:
<https://powerbi.microsoft.com/ru-ru/>
33. OLAP [Электронный ресурс]/ URL:
<https://ru.wikipedia.org/wiki/OLAP>
34. Prognoz Platform [Электронный ресурс]/ URL:
<http://www.prognoz.ru/platform/download>
35. QlikView [Электронный ресурс]/ URL: <http://global.qlik.com/ru/landing/go-sm/qlikview/download-qlikview/>
36. Tableau [Электронный ресурс]/ URL:
<https://www.tableau.com/products/trial>
37. Tibco Spotfire [Электронный ресурс]/ URL:
<https://spotfire.tibco.com/overview>